



**Universidad de Valladolid**  
**Facultad de Ciencias Económicas**  
**y Empresariales**

**Trabajo Fin de Grado**

**Grado** en Marketing e Investigación de Mercados

**Comparativa Económica de las**  
**Comunidades Autónomas españolas en el**  
**siglo XXI: Un análisis multivariante.**

Presentado por:

***Juan Carlos Herrero Cesteros***

Tutelado por:

***M<sup>o</sup> Araceli Rodríguez López***

Valladolid, Marzo de 2023



## **RESUMEN**

El objetivo de este Trabajo de Fin de Grado es comparar las distintas Comunidades Autónomas de España a través de sus variables más relevantes mediante un análisis descriptivo, con el fin de descubrir las características de la economía española del siglo XXI. Según nuestro análisis descriptivo, los mejores años para realizar el estudio del análisis multivariante son 2010, 2015 y 2021.

Para llevar a cabo el análisis multivariante, se utilizarán dos técnicas: primero, el análisis factorial para resumir las variables; y luego, el análisis clúster para agrupar o separar las Comunidades Autónomas. Estas técnicas nos permitirán identificar las características fundamentales de las agrupaciones resultantes y cómo varían en cada año.

Palabras clave: Comunidades Autónomas, Análisis factorial, Análisis clúster, Análisis multivariante.

Códigos de la clasificación JEL: C38, R11.

## **ABSTRACT**

This Final Degree Project is to compare the different Autonomous Communities of Spain through their most relevant variables using a descriptive analysis in order to discover the characteristics of the Spanish economy in the 21st century. Our descriptive analysis allows us to understand that the best years to carry out the multivariate analysis study are 2010, 2015 and 2021.

To perform the multivariate analysis, two techniques will be used: first, factor analysis to summarize the variables; and then, cluster analysis to group or separate the Autonomous Communities. These techniques will allow us to identify the fundamental characteristics of the resulting groups and how they vary in each year.

Keywords: Autonomous Communities of Spain, Factorial analysis, Cluster analysis, Multivariate analysis.

Classification codes JEL: C38, R11.

## ÍNDICE

INTRODUCCIÓN.....	8
1 ANÁLISIS DESCRIPTIVO DE SITUACIÓN DE LAS CCAA.....	9
1.1 Selección de datos.....	9
1.2 Caracterización de la Economía Española de principios del siglo XXI. ....	11
2. ANÁLISIS FACTORIAL Y ANÁLISIS DE CLÚSTER .....	30
2.1 ¿Qué es un Análisis Factorial?.....	30
2.2 Análisis Clúster.....	34
3. ANALISIS MULTIVARIANTE APLICADO A LAS CCAA .....	37
3.1 Elección y tratamiento de las variables .....	37
3.2 Caso 1: año 2010 .....	41
3.3 Caso 2: Año 2015.....	49
3.4 Caso 3: Año 2021.....	57
4. CONCLUSIONES.....	66
5. BIBLIOGRAFÍA .....	68
6. ANEXOS.....	71

## ÍNDICE DE TABLAS Y GRÁFICOS.

Tabla 1.1. Población media 2002-2021 .....	13
Gráfico 1.1. Tasa De Variación Anual de la Población Media por CCAA (2002-2021) .....	14
Tabla 1.2. Edad Media de la Población (2002-2021).....	15
Gráfico 1.2. Tasa de Variación Anual de las CCAA que pierden población media (2002-2021) .....	16
Gráfico 1.3 Tasa de Variación Anual de la Población y la Edad Media (2002-2021).....	17
Tabla 1.3 PIB Medio (2002-2021).....	19
Tabla 1.4 PIB Per cápita Medio (2002-2021) .....	20
Gráfico 1.5 Tasa De Variación Anual Del PIB (2002-2021) .....	21
Gráfico 1.6. Tasa De Variación Anual Del PIB por CCAA (2002-2021).....	21
Gráfico 1.7 Tasa De Paro y Riesgo De Pobreza (2008-2021) .....	22
Gráfico 1.8 Tasa De variación Anual de la Tasa de Paro y Riesgo de Pobreza por CCAA (2008-2022) .....	23
Gráfico 1.9 Rentas y Gastos Medios por CCAA (2008-2021) .....	24
Gráfico 1.10 Tasa de Variación Anual de las Rentas y Gastos Medios (2008-2021).....	25
Gráfico 1.11 Índice de precios de vivienda y consumo (2008-2021) .....	26
Gráfico 1.12 Tasa de Variación Anual de los precios de vivienda e IPC (2008-2021).....	26
Tabla 1.5 Índice de producción industrial media (2008-2021).....	27
Gráfico 1.13 Coste Laboral y Gasto en investigación en porcentaje del PIB por CCAA (2008-2021) .....	28
Tabla 3.1 Variables .....	39
Tabla 3.2 Pertinencia Del Análisis .....	41
Tabla 3.3 Varianza Total Explicada .....	41
Tabla 3.4 Comunalidades .....	42
Tabla 3.5 Matriz De Componentes Rotados .....	42

Tabla 3.6 Determinación Del Número De Clústeres .....	44
Tabla 3.7 Caso De Clúster.....	45
Tabla 3.8 Resumen De Casos .....	46
Tabla 3.9 Índice de Silueta .....	48
Tabla 3.10 Pertinencia Del Análisis .....	49
Tabla 3.11 Varianza Total Explicada .....	50
Tabla 3.12 Comunalidades .....	50
Tabla 3.13 Matriz De Componentes Rotados .....	51
Tabla 3.14 Determinación Del Número De Clúster .....	52
Tabla 3.15 Caso De Clúster .....	53
Tabla 3.16 Resumen De Casos.....	54
Tabla 3.17 Índice De Silueta .....	56
Tabla 3.18 Pertinencia Del Análisis .....	57
Tabla 3.19 Varianza Total Explicada .....	58
Tabla 3.20 Comunalidades .....	58
Tabla 3.21 Matriz De Componentes Rotados .....	59
Tabla 3.22 Determinación Del Número De Clúster .....	61
Tabla 3.23 Caso De Clúster.....	61
Tabla 3.24 Resumen De Casos.....	62
Tabla 3.25 Índice De Silueta .....	65
Tabla A.1 Tasa De Paro.....	71
Gráfico A.1 Tasa De Paro .....	71
Tabla A.2 Tasa De Variación Anual Del IPC.....	72
Gráfico A.2 Tasa De Variación Anual Del IPC .....	72
Gráfico A.3 Índice De Vivienda: Inversión Y Precios. ....	73
Tabla A.3 Deuda Pública En Porcentaje Del PIB .....	73
Gráfico A.4 Deuda Pública En Porcentaje Del PIB .....	74

Tabla A.4 Crecimiento Vegetativo: Natalidad Y Mortalidad.....	74
Gráfico A.5 Crecimiento Vegetativo: Natalidad Y Mortalidad .....	75
Tabla A.5 Tasa De Natalidad: española Y Extranjera .....	75
Tabla A.6 Correlaciones Pearson 2010.....	76
Tabla A.7 Correlaciones Pearson 2015.....	77
Tabla A.8 Correlaciones Pearson 2021.....	78
Tabla A.9 Conjunto De Datos 2010 .....	79
Tabla A.10 Conjunto De Datos 2015 .....	80
Tabla A.11 Conjunto De Datos 2021 .....	81
Tabla A.12 IPC Medio CCAA (2008-2021).....	82

## **INTRODUCCIÓN**

Las Comunidades Autónomas españolas presentan grandes diferencias en términos económicos, sociales, demográficos y territoriales. Esto no es solo un problema estructural, sino que se observan problemas coyunturales en distintos horizontes temporales en este siglo, como por ejemplo durante la crisis del 2008, la etapa de recuperación del 2015 o la pandemia Covid-19.

En este trabajo se lleva a cabo una comparativa entre las distintas Comunidades Autónomas de España a través de diferentes análisis estadísticos multivariantes. El objetivo es mostrar las diferencias existentes entre ellas y cómo estas han ido evolucionando a lo largo del siglo XXI. Para ello, se utiliza la siguiente metodología: en primer lugar, se realiza un análisis descriptivo de las variables más relevantes separadas en sociodemográficas y macroeconómicas en cada Comunidad Autónoma, excluyendo Ceuta y Melilla. Posteriormente, se procede a realizar un análisis factorial para resumir la información de las variables en dimensiones o factores. Esta técnica nos permitirá guardar las puntuaciones de los factores para luego agrupar las Comunidades en función de sus similitudes o diferencias mediante el análisis clúster. Por último, se observan las diferencias producidas en las distintas etapas analizadas (2010-2015-2021). Este enfoque nos permitirá comprender mejor la compleja realidad del territorio español y las diferentes situaciones económicas, sociales y culturales que caracterizan a cada Comunidad Autónoma.

En conclusión, el objetivo principal de este trabajo es comprender e interpretar las diferencias que existen entre las Comunidades Autónomas, estableciendo correlaciones entre variables para observar cómo estas afectan a cada territorio en particular. Además, se busca analizar cómo varían estas correlaciones en cada horizonte temporal, con el fin de tener una comprensión más completa de la dinámica territorial a lo largo del tiempo.

# **1 ANÁLISIS DESCRIPTIVO DE SITUACIÓN DE LAS CCAA**

## **1.1 Selección de datos.**

Para este análisis descriptivo y los posteriores análisis estadísticos, se ha excluido Ceuta y Melilla de la muestra de Comunidades Autónomas. Esta decisión se ha tomado debido a las dificultades para encontrar bases de datos fiables para estas dos regiones, así como por las significativas diferencias en su estructura y características con respecto al resto de las comunidades autónomas. Por tanto, la muestra final seleccionada para el análisis se compone de un total de 17 Comunidades Autónomas.

En cuanto a las variables analizadas, pueden ser agrupadas en sociodemográficas y macroeconómicas.

Las variables sociodemográficas son aquellas que nos van a permitir explicar las características de la sociedad. Para ello, voy a utilizar variables relacionadas con la población, como los flujos migratorios, la edad media que tiene la población y por último el saldo vegetativo compuesto por la diferencia entre mortalidad y natalidad.

La “Población” es un buen indicador sociodemográfico que nos permite explicar variables tanto económicas como sociales (PIB per cápita, Tasa de Paro, Rentas medias). Además, la población es interesante compararla con las demás variables sociodemográficas para obtener un significado más amplio, ya que las variaciones de la población pueden corresponder a fenómenos demográficos tales como el envejecimiento, el saldo vegetativo o los movimientos migratorios.

La variable, “Edad media de la población” nos permite identificar aquellas Comunidades Autónomas con poblaciones más envejecidas lo cual puede afectar al saldo vegetativo y puede influir sobre algunas variables macroeconómicas como, por ejemplo, a la tasa de actividad.

Por otro lado, las variables macroeconómicas son aquellas que explican el comportamiento o la situación de la economía. Para ello, voy a analizar variables como el PIB a precios de mercado, la tasa de paro, el riesgo de pobreza y los índices de los precios de vivienda y el IPC (índice de precios al consumidor). En relación al grado de industrialización de las CCAA, voy a estudiar, el Índice de producción industrial, el coste laboral y el gasto por investigación en porcentaje

del PIB. Finalmente, también se analizarán, características económicas de las unidades familiares como la renta media por hogar, el gasto medio por persona y la renta media por unidad de consumo.

El “PIB” o la “tasa de paro”, determinarán la situación coyuntural de la economía. La variable “riesgo de pobreza” nos indica qué porcentaje de la población está viviendo en unas condiciones precarias. Los dos índices de precios (IPC y el IPV) nos permitirán determinar las diferencias de precios que existen en las CCAA.

Los componentes de la industrialización miden la actividad productiva industrial. Se ha incluido el “Coste laboral industrial” y el “Gasto en investigación y desarrollo en porcentaje del PIB” para observar las diferencias que existen entre Comunidades Autónomas y determinar cuáles están más industrializadas. Por último, se ha incluido el “IPI”, índice de producción industrial que incluye todas aquellas actividades como las extractivas, manufactureras y de producción/distribución<sup>1</sup>.

La variable "Renta Media por unidad de consumo" nos proporciona información sobre la renta media de las personas, mientras que la variable "Renta media por hogar" nos permite analizar cómo se distribuyen los recursos económicos al compartir los gastos entre los miembros de una misma familia. Ambas variables son útiles para observar las rentas disponibles, por lo tanto, nos sirven para comparar las diferencias económicas entre las distintas Comunidades Autónomas. Estas variables están relacionadas con el “Gasto medio por persona” ya que el gasto depende de la renta disponible. Así, podemos determinar cuáles son las regiones más prósperas y comprender las razones detrás de estas diferencias.

Todas las variables utilizadas en el análisis descriptivo han sido obtenidas del INE (Instituto Nacional de Estadística) ya que ofrece una gran cantidad de bases de datos segmentadas por distintas variables como pueden ser económicas, sociales, demográficas que permiten separarlas por años y territorios. Además, es posible consultar toda la base de datos utilizada para este trabajo a través de este enlace:

---

<sup>1</sup> La metodología utilizada para la obtención del “Índice de Producción industrial” por parte del INE, así como sus ponderaciones en cada sector industrial se puede consultar en este enlace: [https://www.ine.es/prensa/jpi\\_prensa.htm](https://www.ine.es/prensa/jpi_prensa.htm)

<https://drive.google.com/drive/folders/1pgfyJrm7hIGMpqyhfnfx1IQbSQJYjssBZ?usp=sharing>

Por último, es interesante estudiar la evolución de cada variable así como los cambios anuales, mediante tasas de variación.

## **1.2 Caracterización de la Economía Española de principios del siglo XXI.**

La economía española del siglo XXI está marcada lógicamente, por fases de recesión (2008-2014, 2020-Actualidad) y fases de expansión o recuperación (2000-2007, 2015-2019).

Antes de la crisis del 2008 denominada como “La Gran recesión”, la economía española estuvo marcada por el ingreso en la Comunidad Económica Europea (CEE) lo que permitió incrementar el PIB per cápita en relación con otros países de la UE-15 hasta prácticamente alcanzar los niveles de Italia y Francia. Desde 1986-2007 se produjo un crecimiento de la tasa media anual acumulativa del 3,4%. También, se produjeron desequilibrios macroeconómicos como el aumento de los precios (42%) con la consecuente pérdida de competitividad y un déficit de la balanza de saldos y un incremento de la deuda pública insostenible<sup>2</sup>

Entre 2008-2013 se produjo un retroceso de la economía de tal forma que se redujo el PIB real en media 1,8 puntos por año y disminuyó el empleo, Desde 2010 han perdido peso los sectores de la agricultura, construcción e industria, y se ha producido una terciarización de la economía con turismo, sanidad y educación ganando peso sobre el PIB, reduciendo la productividad.

La Tasa de Paro (Tabla A.1 del Anexo) es un variable macroeconómica que en nuestro país se caracteriza por ser bastante alta en comparación con otros países desarrollados. Los valores más elevados de la tasa de paro se produjeron durante la crisis superando el 20% y alcanzando el máximo en 2013 con un 25,73%. El mínimo de este siglo se produce en 2006 con un 8,26% cercano a los demás países europeos. Sin embargo, el último dato de 2021 sitúa la tasa en un 13,33%, muy lejos de la media europea.

El desempleo provoca un ciclo de retroceso económico ya que las personas que

---

<sup>2</sup> Laborda, Á. (2016).

se quedan sin trabajo ven reducidos sus gastos y su consumo, incidiendo en un círculo donde las empresas disminuyen su producción porque caen sus ventas y, en consecuencia, disminuye la producción del país.

Debemos referirnos, en términos generales a la gran subida del IPC (Tabla A.2 del Anexo) producida en 2021, con un 6.5% de variación anual debido principalmente al conflicto en Ucrania y la pandemia covid-19. El resto de los años no hay diferencias significativas, excepto en el 2007 que la tasa de variación anual aumenta en 4,2%. Por el contrario, el IPV (índice de vivienda)<sup>3</sup> aumentó hasta 2007-2008 para caer significativamente hasta 2013-2014 donde comienza una tendencia alcista muy moderada (gráfico A.3 del Anexo).<sup>4</sup>

En cuanto a la deuda pública, en porcentaje del PIB, (Tabla A.3 del Anexo) ha aumentado desde la crisis del 2008, alcanzando en 2013, un 100,5%. Esto significa que toda la producción generada en España durante todo un año es igual a la deuda que tenemos. En 2022, sigue estando por encima de la producción nacional con un 116,1%. Una deuda pública elevada puede generar problemas de financiación de las políticas públicas, pérdidas de competitividad y de credibilidad.

Para comprender el panorama sociodemográfico actual, es necesario examinar detenidamente la evolución reciente de las principales variables demográficas y sociales que comento a continuación.

La tabla 1.1 muestra la población media de cada Comunidad Autónoma entre los años 2002 y 2022, ordenadas de mayor a menor. Para analizar su distribución, se ha asignado una situación en función de los cuartiles a los que pertenecen. La variable "población" ha sido seleccionada como la primera en ser analizada, ya que se pretende comprobar si está relacionada con el resto de las variables. Además, se pretende demostrar su incidencia sobre las variables sociodemográficas y si también está relacionada con algunas variables macroeconómicas, como el PIB.

En primera instancia, podemos observar la gran variabilidad entre Comunidades,

---

<sup>3</sup> El IPV, índice de vivienda mide los precios de compraventa y su evolución a lo largo del tiempo, se puede consultar más información y su metodología en este enlace: [https://www.ine.es/prensa/ipv\\_prensa.htm](https://www.ine.es/prensa/ipv_prensa.htm)

<sup>4</sup> Hernández, P. (2018)

existiendo una diferencia muy relevante entre Madrid y Cantabria dado que tienen una extensión muy similar pero una población media muy distinta, por poner un ejemplo.

**Tabla 1.1. Población media 2002-2021**

Situación CCAA		Media
Alta	Andalucía	8.207.648,8
	Cataluña	7.365.246,5
	Madrid	6.317.210,6
	Comunidad Valenciana	4.914.575,5
Estándar	Galicia	2.748.454,4
	Castilla y León	2.487.796,6
	País Vasco	2.170.822,8
	Canarias	2.067.581,7
	Castilla - La Mancha	2.008.865,2
	Murcia	1.425.583,9
	Aragón	1.307.347,7
	Extremadura	1.086.787,3
	Baleares	1.075.341,1
	Asturias	1.060.023,0
Baja	Navarra	627.120,8
	Cantabria	577.876,1
	Rioja	312.430,4
Cuartiles	Q1	1.067.682,0
	Mediana	2.008.865,2
	Q3	3.831.514,9

*Fuente: Elaboración propia a partir de datos del INE.*

En general, podemos establecer que, en término medio, las CCAA con territorios de menor superficie son aquellas que tienen una población menor, lógicamente. Sin embargo, hay que matizar porque algunas, como Madrid, siendo territorialmente pequeña, es una Comunidad muy poblada. Esto nos indicaría que tiene una densidad de habitantes por kilómetro cuadrado<sup>5</sup> muy elevada.

Sin embargo, esta variable nos da poca información sin tener en cuenta otra serie de factores, como muestra el gráfico 1.1, elaborado a partir de los valores de 2002 y 2021 para comprobar si ha aumentado o disminuido la población entre esas dos fechas.

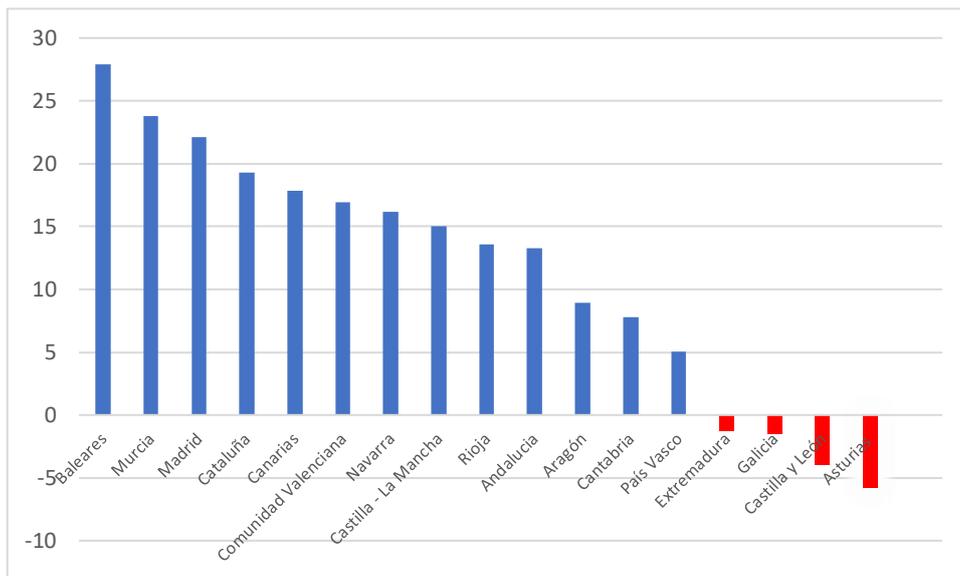
En azul se presentan aquellas CCAA, cuya población ha crecido y en rojo,

<sup>5</sup> La densidad nos indica la cantidad de habitantes por kilómetro cuadrado que conforma un territorio, destacando Castilla y León (25.34) y Madrid (841,15) como las Comunidades con menor densidad y mayor, respectivamente [https://es.wikipedia.org/wiki/Anexo:Comunidades\\_y\\_ciudades\\_aut%C3%B3nomas\\_de\\_Espa%C3%B1a](https://es.wikipedia.org/wiki/Anexo:Comunidades_y_ciudades_aut%C3%B3nomas_de_Espa%C3%B1a)

aquellas otras que han perdido (ordenadas de mayor a menor crecimiento).

La primera comparación que podemos realizar es que no podemos determinar que las Comunidades que tienen baja población al mismo tiempo la pierdan, sin embargo, sí que podemos decir que las Comunidades que más crecen son aquellas que tienen más población.

**Gráfico 1.1. Tasa De Variación Anual de la Población Media por CCAA (2002-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

Para buscar conclusiones que determinen la pérdida de población de estas Comunidades, en la tabla 1.2 podemos observar cómo Asturias, Castilla y León y Galicia son las Comunidades más envejecidas lo que puede explicar la pérdida de población. Además, aquellas que más crecen, como Baleares y Murcia, son las que tienen la población más joven.

Parece que “Población” y “Edad media de la población” están correlacionadas, ya que, si la población es más joven, hay menos mortalidad y más natalidad lo que permite que el saldo vegetativo sea mayor que en aquellos territorios con población más envejecida.

**Tabla 1.2. Edad Media de la Población (2002-2021)**

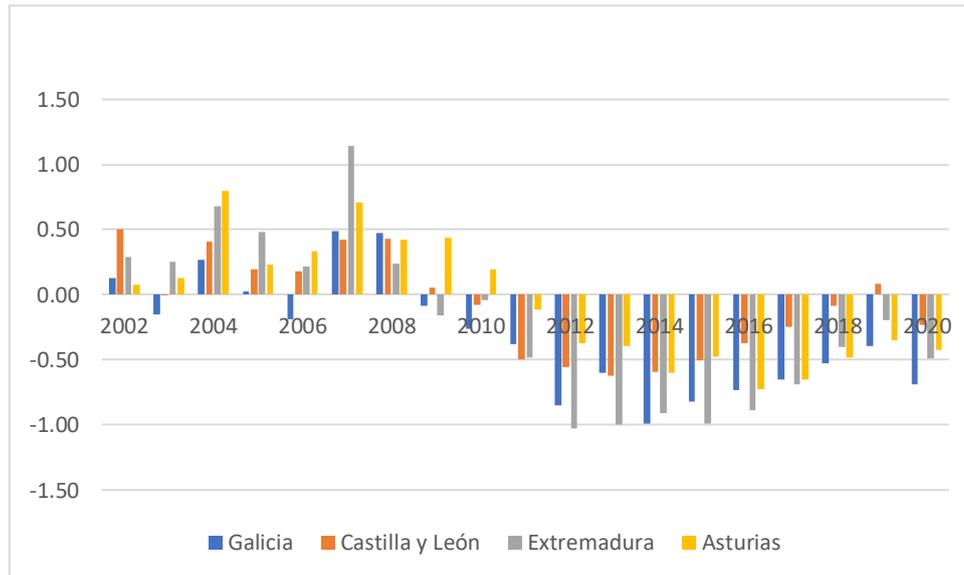
Situación CCAA		Media
Envejecida	Asturias	46,3
	Castilla y León	45,6
	Galicia	45,3
	País Vasco	43,8
Estándar	Cantabria	43,7
	Aragón	43,6
	Rioja	42,8
	Extremadura	42,3
	Navarra	42,0
	Cataluña	41,5
	Comunidad Valenciana	41,5
	Castilla - La Mancha	41,5
Jóven	Madrid	40,6
	Baleares	40,0
	Canarias	39,9
	Andalucía	39,8
Cuartiles	Murcia	38,8
	Q1	40,3
	Mediana	42,0
	Q3	43,8

*Fuente: Elaboración propia a partir de datos del INE.*

En el gráfico 1.2, se presenta la relación entre la variable de población y la edad media de la población en las distintas Comunidades Autónomas medidas en tasa de variación anual. Al analizar los resultados, podemos observar que las cuatro comunidades que han experimentado una pérdida de población son precisamente aquellas que presentan una edad media de población muy alta. Esto refuerza la hipótesis de que existe una relación entre la disminución de la población y el envejecimiento demográfico. Además, se puede determinar si la pérdida de población fue simultánea y en qué momento ocurrió.

Como podemos observar desde 2002 hasta prácticamente 2008 hay crecimiento de la población. El problema parece comenzar entre 2010, donde Extremadura, Castilla y León y Galicia comienzan a perder población. En 2011 también se incorpora Asturias. Extremadura y Galicia son las que más población en término medio pierden durante estos años.

**Gráfico 1.2. Tasa de Variación Anual de las CCAA que pierden población media (2002-2021)**

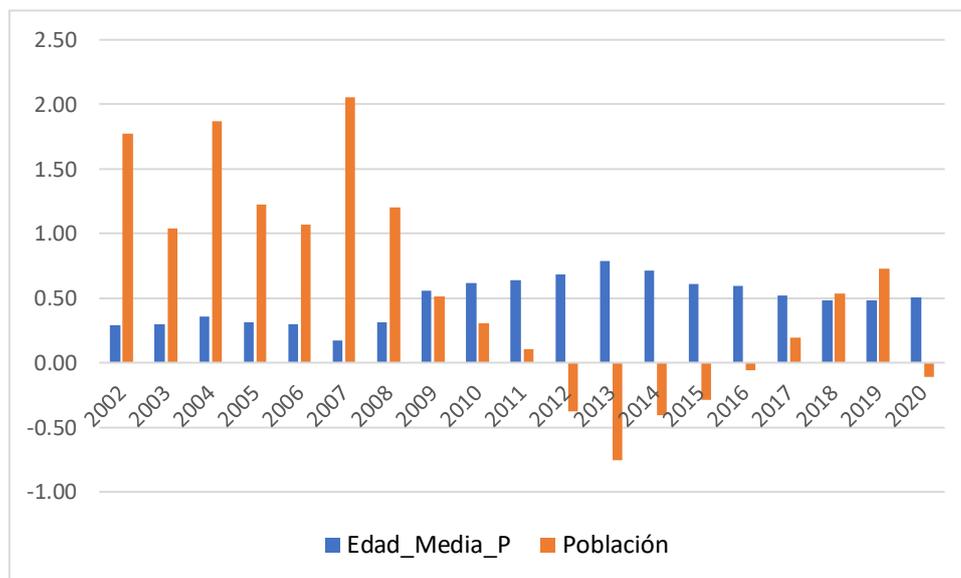


*Fuente: Elaboración propia a partir de datos del INE.*

Con el gráfico 1.3 se ha buscado demostrar si la relación entre la variable de población y la edad media de la población es similar en todas las Comunidades Autónomas, independientemente de si experimentan pérdidas de población o no. Al analizar el gráfico, podemos observar que existe una relación inversamente proporcional entre ambas variables. Esto significa que, a medida que aumenta la edad media disminuye el crecimiento de la población.

Desde 2008 las tasas de variación anual de la población media comienzan a disminuir de tal forma que en 2012-2016 son negativas. Todo parece indicar, que, en este horizonte temporal, existe un problema común en España por el cual la población disminuye con respecto a años anteriores.

**Gráfico 1.3 Tasa de Variación Anual de la Población y la Edad Media (2002-2021)**

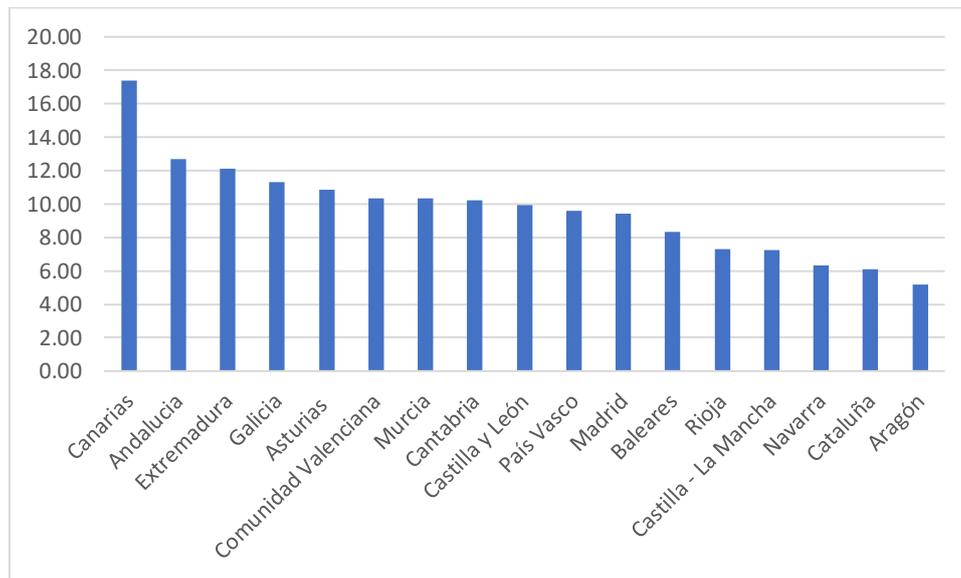


*Fuente: Elaboración propia a partir de datos del INE.*

En los primeros años de la década del 2000, la población envejece muy poco a poco, pero entre 2008 y 2015, la población envejece muchísimo más ya que la natalidad en épocas de crisis disminuye al haber menos seguridad y capacidad monetaria en las familias por lo que deciden no tener hijos. Además, la disminución de la inmigración y el aumento de la emigración en España, sobre todo de jóvenes, puede ser una explicación adicional.

En el gráfico 1.4 se observa el cambio en el envejecimiento de la población en las distintas Comunidades Autónomas entre 2002 y 2021, permitiendo identificar aquellas que más han experimentado este fenómeno. Es importante destacar que en todas las Comunidades Autónomas ha ocurrido un envejecimiento de la población, lo que indica una mayor esperanza de vida al mismo tiempo que disminuye la natalidad y la mortalidad.

**Gráfico 1.4 Tasa de Variación Anual de la edad media de la población por CCAA (2002-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

Comparativamente, la población que más se envejece es Canarias, a pesar de que tenga una de las poblaciones más jóvenes. Sin embargo, si podemos observar la relación entre el mayor incremento del envejecimiento de la población y la pérdida de población. Galicia, Extremadura y Asturias, que pierden población desde 2010 a 2021, posiblemente deban esa pérdida a que su población está envejecida y además sigue envejeciéndose más rápido que en las demás comunidades.

Esta pérdida de población se puede explicar desde el punto de vista del saldo migratorio (Tabla A.4 del Anexo). Antes de la crisis de 2008, se producía una llegada muy importante de inmigrantes, que se vio interrumpida a partir de 2007, por la crisis económica. Además, comienza un proceso de crecimiento de la emigración, principalmente desde 2013, lo que repercute en un saldo migratorio negativo desde 2010 hasta 2015, es decir, una vez finalizada la crisis. Desde entonces, disminuye de nuevo la emigración y aumenta la inmigración, con lo que se produce un saldo migratorio positivo, y, por tanto, aumenta la población en España como hemos podido constatar.<sup>6</sup> Los inmigrantes favorecen la natalidad ya que su tasa de natalidad frente a la población española es bastante

<sup>6</sup> Hernández Vera, Pablo (2020)

superior y con ello el saldo vegetativo, con el consecuente incremento de la población (Tabla A.5 del Anexo).

A continuación, se presenta un análisis descriptivo de algunas de las principales variables macroeconómicas.

En la tabla 1.3 se recoge el PIB en media entre los años 2002-2021 (ordenado de mayor a menor). Podemos observar la gran diferencia que existe entre Comunidades como Madrid y Cantabria que territorialmente son prácticamente similares. Es importante tener en cuenta que, por lo general, una mayor población suele estar correlacionada con un mayor Producto Interno Bruto (PIB). Debido a esto, simplemente comparar el PIB total de ambas regiones no es suficiente para obtener conclusiones significativas. Por lo tanto, resulta necesario comparar el PIB per cápita (es decir, el PIB dividido por la población) para obtener una visión más precisa de cómo se distribuye la riqueza en cada región. Estos datos se encuentran presentados en la tabla 1.4.

**Tabla 1.3 PIB Medio (2002-2021)**

Situación CCAA		Media
Alto	Cataluña	197.935.641
	Madrid	195.656.707
	Andalucía	141.218.852
	Comunidad Valenciana	98.963.401
Estándar	País Vasco	62.946.681
	Galicia	54.673.538
	Castilla y León	52.853.000
	Canarias	39.835.040
	Castilla - La Mancha	36.661.514
	Aragón	32.599.402
	Murcia	27.198.312
	Baleares	26.607.255
	Asturias	21.397.586
Bajo	Navarra	17.584.248
	Extremadura	17.361.354
	Cantabria	12.281.826
	Rioja	7.672.037
Cuartiles	Q1	19.490.917
	Mediana	36.661.514
	Q3	80.955.041

*Fuente: Elaboración propia a partir de datos del INE.*

Cataluña y Madrid son las Comunidades que más aportan en producto interior bruto y que La Rioja y Cantabria son las que menos, existiendo mucha diferencia.

A partir de la tabla 1.4 sí que podemos determinar que Comunidades son más

ricas en términos económicos si las comparamos con la cantidad de población que tienen. La primera gran diferencia que podemos observar es que Andalucía a pesar de tener un PIB muy alto, si lo comparamos con la población es de los más bajos. Esto nos indica que Andalucía crea mucha riqueza ya que es muy grande territorialmente y además está muy poblada, pero sus habitantes son más pobres en términos comparativos.

**Tabla 1.4 PIB Per Cápita Medio (2002-2021)**

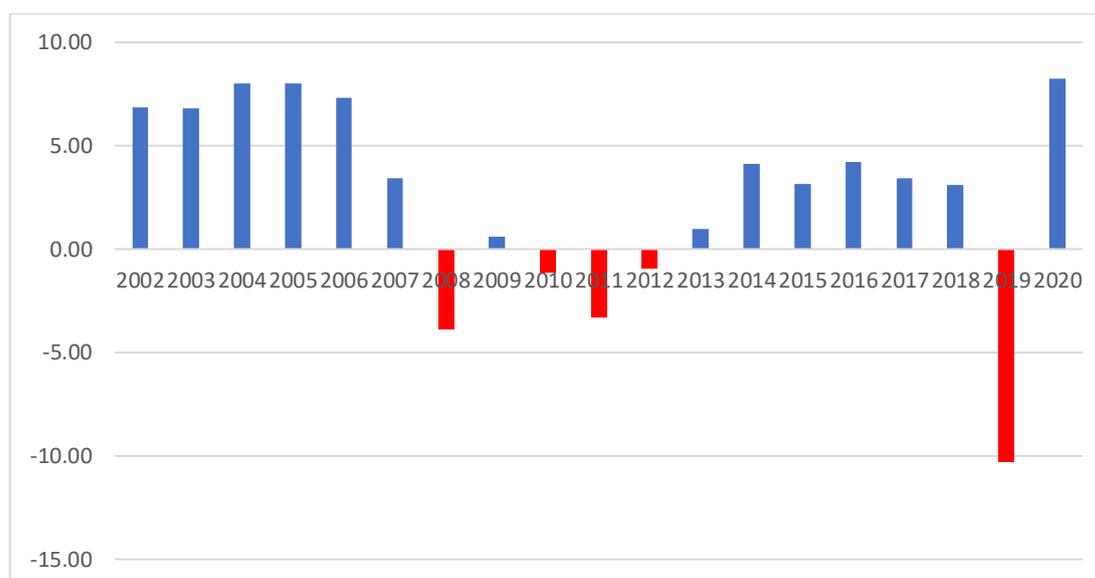
Situación CCAA		Media
Alto	Madrid	30,97
	País Vasco	29,00
	Navarra	28,04
	Cataluña	26,87
Estándar	Aragón	24,94
	Baleares	24,74
	Rioja	24,56
	Cantabria	21,25
	Castilla y León	21,24
	Asturias	20,19
	Comunidad Valenciana	20,14
	Galicia	19,89
	Canarias	19,27
Murcia	19,08	
Bajo	Castilla - La Mancha	18,25
	Andalucía	17,21
	Extremadura	15,97
Cuartiles	Q1	19
	Mediana	21
	Q3	26

*Fuente: Elaboración propia a partir de datos del INE.*

En el caso contrario, Madrid y Cataluña generan mucha riqueza y además sus habitantes, en término medio, poseen mayor riqueza.

En el gráfico 1.5 podemos observar que desde 2008-2012 y en tiempos de la pandemia, desde 2019, las tasas anuales del PIB son negativas, lo que refleja las dos crisis y la recesión económica producida.

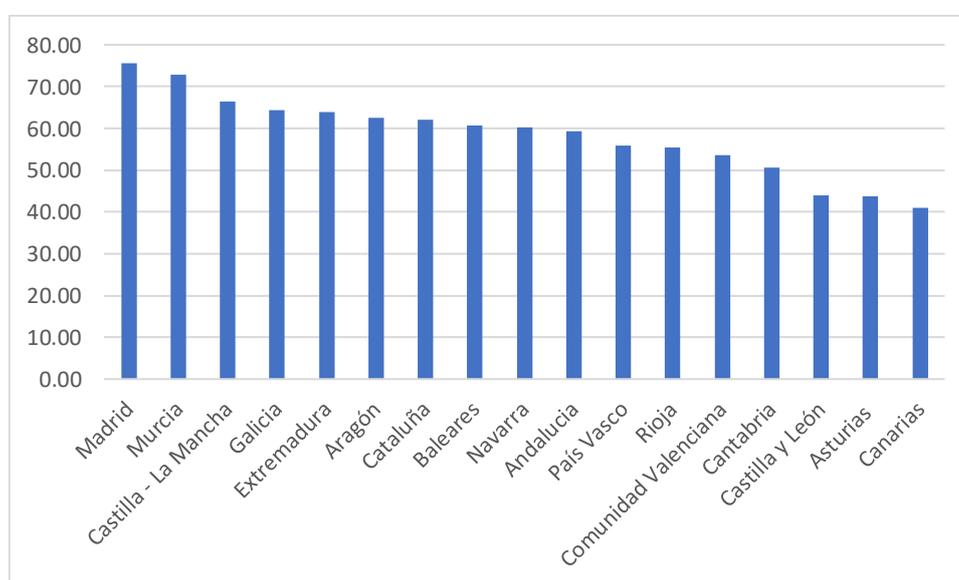
**Gráfico 1.5 Tasa De Variación Anual Del PIB (2002-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

En el gráfico 1.6 podemos observar que todas las Comunidades han visto incrementado su PIB desde el 2002, destacando Madrid y Murcia como las que más crecimiento han conseguido en el periodo analizado.

**Gráfico 1.6. Tasa De Variación Anual Del PIB por CCAA (2002-2021)**

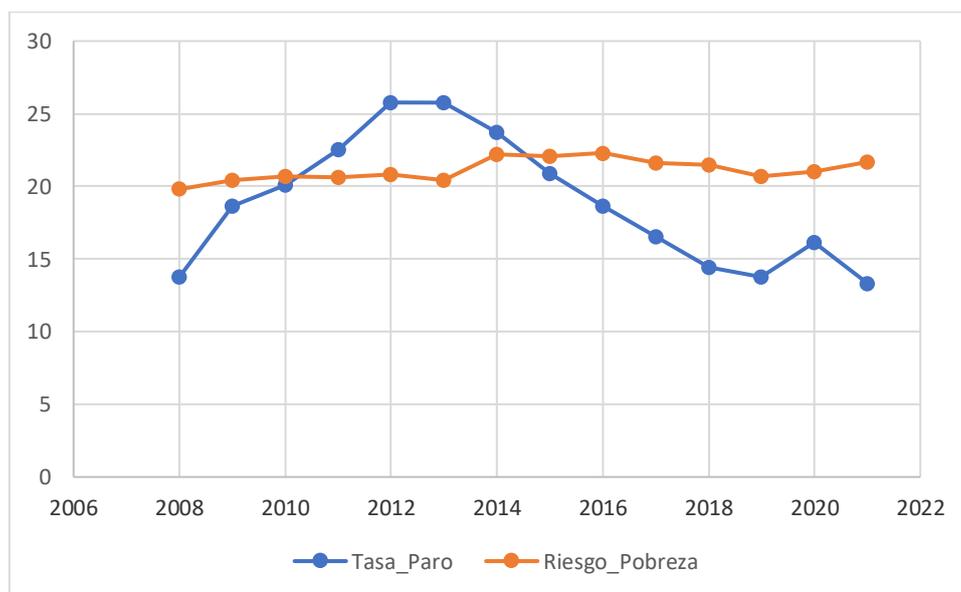


*Fuente: Elaboración propia a partir de datos del INE.*

Como ya hemos comentado anteriormente, otro indicador económico muy importante es la “Tasa de Paro” que nos mide el número de parados en función de las personas que están en condiciones y disponibles para trabajar (población

activa) pero no consiguen trabajo. En el gráfico 1.7 podemos observar cómo esta tasa es más alta en las épocas de crisis (2008-2014) y finalmente en (2019-2020), lo que nos explica una clara relación con el PIB. Cuando la Tasa de Paro aumenta, el PIB se ve reducido en variación anual incluso llegando a ser negativa.<sup>7</sup>

**Gráfico 1.7 Tasa De Paro y Riesgo De Pobreza (2008-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

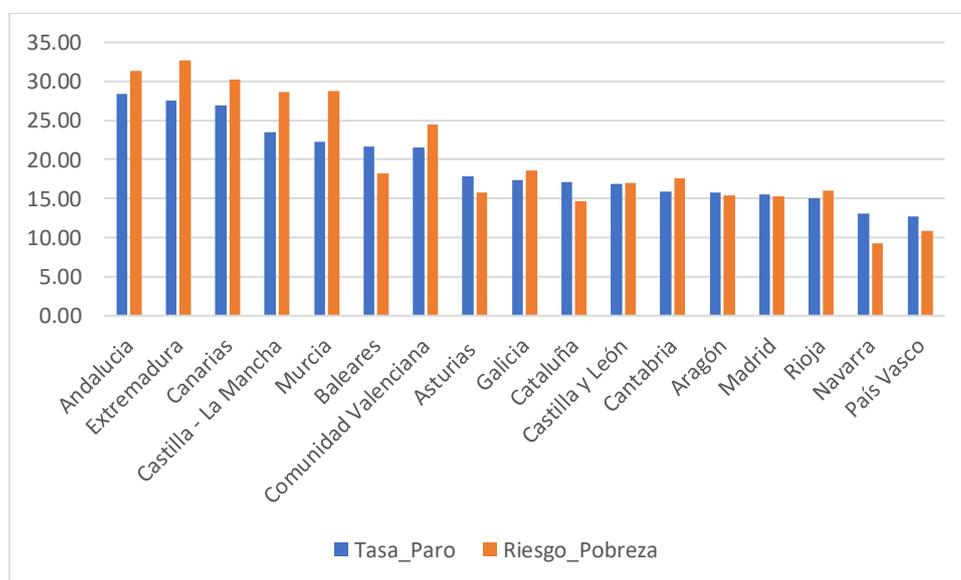
En cuanto al “Riesgo de Pobreza”<sup>8</sup>, sorprendentemente a pesar de que en todas las Comunidades han aumentado su PIB desde 2002, el riesgo de pobreza se ha mantenido e incluso aumentado desde el 2013, llegando en 2014 a situarse por encima de la Tasa de paro. Lo que nos indica es que a pesar de que las personas tengan empleo, su renta no es suficiente para sustentar una vida digna. Tenemos que tener en cuenta que el criterio por el que se establece este umbral es en función de ingresos por unidad de consumo. Si estos aumentan, el umbral por el que se establece este riesgo también aumenta, necesitando más rentas para no alcanzarlo.

<sup>7</sup> Okun, A. M. (1975). “Equality and efficiency”.

<sup>8</sup> Se considera que una persona está en riesgo de pobreza si la renta total está por debajo del umbral de pobreza (60% por debajo de la mediana de renta disponible). Puede consultarse la metodología del INE en este enlace: [https://www.ine.es/ss/Satellite?L=es\\_ES&c=INESeccion\\_C&cid=1259925455948&p=%5C&pagename=ProductosYServicios%2FPYSLayout&param1=PYSDetalle&param3=1259924822888](https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259925455948&p=%5C&pagename=ProductosYServicios%2FPYSLayout&param1=PYSDetalle&param3=1259924822888)

En el gráfico 1.8 podemos observar la relación que tienen la tasa de paro y el riesgo de pobreza. Las CCAA están ordenadas de mayor a menor en función de la Tasa de Paro. En término general, podemos confirmar que un aumento de la tasa de paro implica un aumento del riesgo de pobreza, destacando Andalucía y Extremadura como las Comunidades con el desempleo y el riesgo de pobreza más alto. Si volvemos a la tabla 1.4 también podemos observar como el PIB per cápita en estas Comunidades también es el más bajo, lo que explica que existe una relación entre estas variables: alta tasa de paro implica un bajo PIB per cápita lo que aumenta el Riesgo de pobreza.

**Gráfico 1.8 Tasa De variación Anual de la Tasa de Paro y Riesgo de Pobreza por CCAA (2008-2022)**

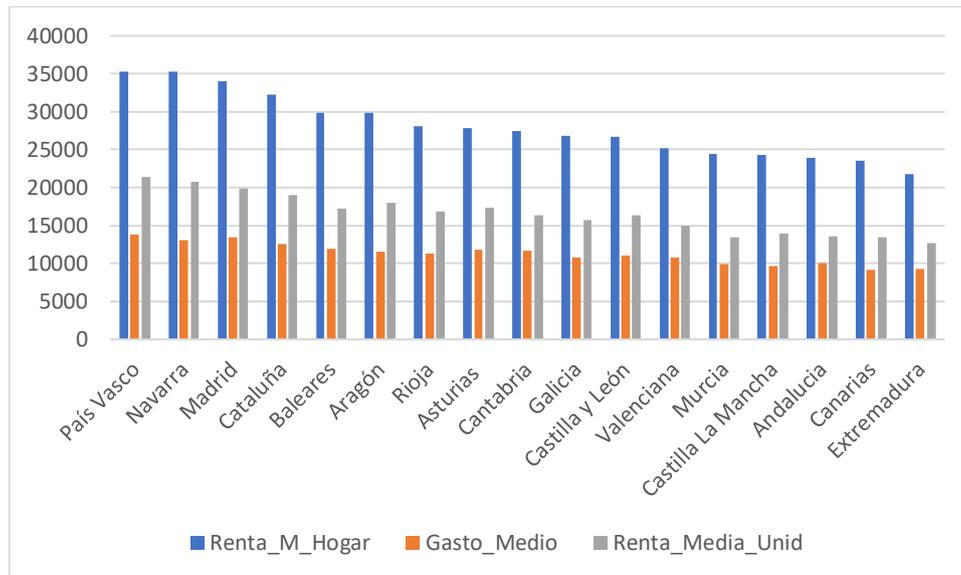


*Fuente: Elaboración propia a partir de datos del INE.*

En el caso contrario, tenemos a Navarra y el País Vasco con el desempleo y riesgo de pobreza más bajo.

En el gráfico 1.9 se representan rentas y gastos medios, ordenados por renta media del hogar de mayor a menor. Podemos observar como una mayor renta media del hogar implica una mayor renta media en unidad de consumo y también un mayor gasto medio. Tiene sentido que, si las familias obtienen más renta, tienen más para consumir, aumentando el gasto y produciendo en la economía un círculo donde las empresas producen más porque los habitantes aumentan su consumo. Esto parece traducirse en una mayor riqueza de esa Comunidad.

**Gráfico 1.9 Rentas y Gastos Medios por CCAA (2008-2021)**



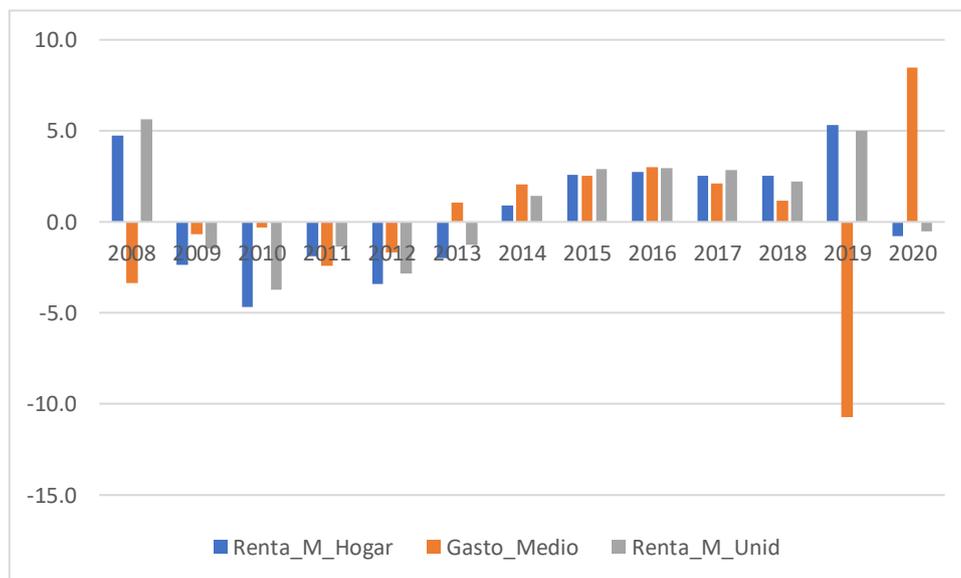
*Fuente: Elaboración propia a partir de datos del INE.*

Podemos observar grandes similitudes con lo comentado anteriormente, ya que también destacan País Vasco y Navarra como las Comunidades con mayores rentas medias y en caso contrario, Extremadura y Andalucía.

Cuando una región tiene una alta tasa de desempleo, suele tener también una renta y un gasto medio más bajo, lo que a su vez se asocia con un mayor riesgo de pobreza y un menor PIB per cápita. En consecuencia, es importante considerar la interrelación de estas variables para entender las dinámicas económicas y sociales en una determinada región. Toda esta relación entre las variables se puede explicar por la “Ley de Okun”.

Con el gráfico 1.10 se pretende demostrar si estas variables (rentas y gastos medios) están relacionadas. Efectivamente, su relación es directamente proporcional en términos generales (cuando una aumenta, lo hace la otra).

**Gráfico 1.10 Tasa de Variación Anual de las Rentas y Gastos Medios (2008-2021)**

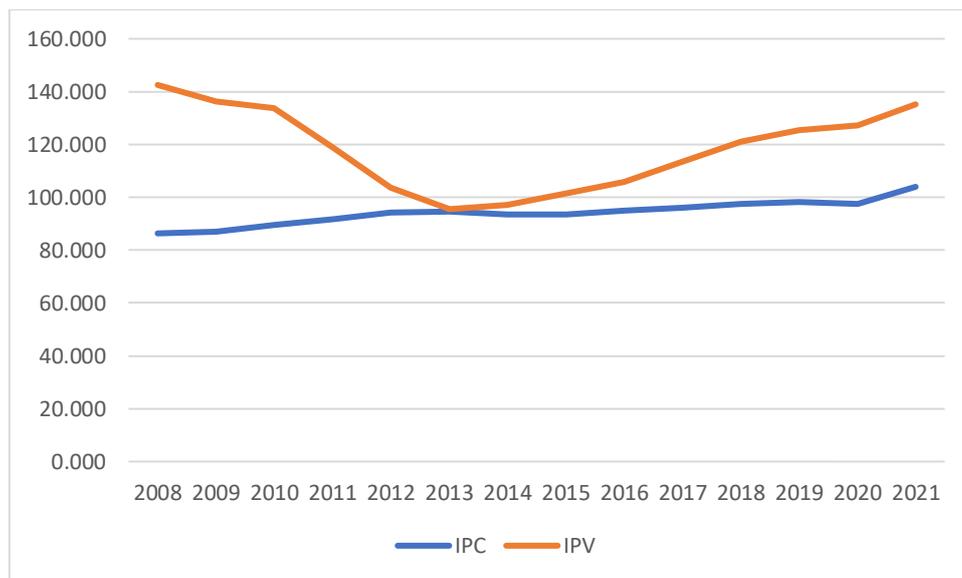


*Fuente: Elaboración propia a partir de datos del INE.*

La recuperación comienza en 2013 (solo en Gasto medio) porque aumenta la confianza y seguridad de la población, con un mayor gasto y un menor ahorro ya que las rentas medias no aumentan. Entre 2014 y 2018 las rentas y los gastos vuelven a crecer dado que estamos primero en fase de recuperación y después en fase expansiva. Esta etapa finaliza en 2019, con la pandemia del Covid-19 en la que, debido a las medidas excepcionales adoptadas, como el confinamiento general de la población, disminuye el gasto medio.

El gráfico 1.11 muestra el índice de precios al consumidor (IPC) y el índice de precios de la vivienda (IPV). Ambos indicadores tienen un efecto directo en los precios y se representan juntos para explorar si están o no relacionados. Al analizar la evolución de ambas variables a lo largo del tiempo, se puede determinar que si existe relación entre ellas y cómo se han visto afectadas en diferentes periodos. En resumen, este gráfico busca analizar la relación entre dos indicadores clave que influyen en los precios, lo que puede proporcionar información valiosa para entender la evolución de la economía.

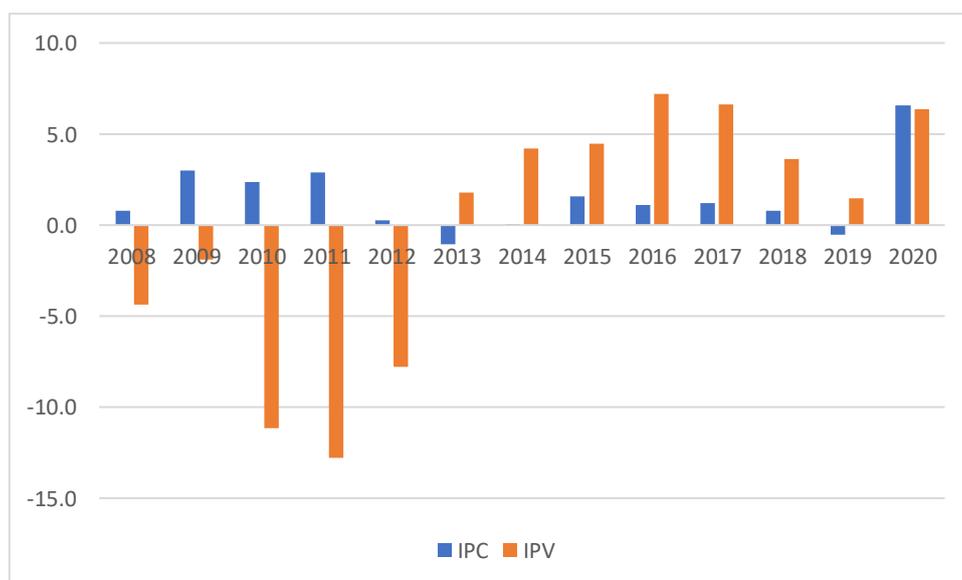
**Gráfico 1.11 Índice de precios de vivienda y consumo (2008-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

Hasta 2013, el IPV disminuye con la crisis del sector de la construcción para luego volver a aumentar. En cuanto al IPC, sigue una tendencia creciente moderada. En la (Tabla A.12 del Anexo) podemos observar cómo no hay diferencias significativas entre las distintas Comunidades.

**Gráfico 1.12 Tasa de Variación Anual de los precios de vivienda e IPC (2008-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

Las tasas de variación, representadas en el gráfico 1.12 parecen mostrar poca

relación entre ambas variables. Desde 2008 hasta 2013, su relación es inversamente proporcional: a medida que aumenta el IPC, el IPV disminuye. Sin embargo, este patrón no se debe a la evolución de los precios, sino a cómo la crisis afectó a diferentes sectores, siendo el de la construcción el más afectado. La caída del sector de la construcción se refleja en el IPV hasta el año 2012, cuando comienza a estabilizarse. En conclusión, aunque existe una aparente relación inversa entre el IPC y el IPV, esta no es la causa directa de la evolución de ambas variables en el periodo analizado, sino que es un reflejo de una situación económica más compleja y de cómo ésta afectó a distintos sectores.

Por último, se muestra el componente de industrialización, medido por el índice de producción industrial (IPI), que se presenta en la tabla 1.5. Vuelven a destacar Madrid y País Vasco con un IPI muy elevado, señal de un alto grado de industrialización. Este es un resultado que coincide con los hallazgos mostrados en las tablas 1.3 y 1.4 (PIB y PIB per cápita), donde ambas Comunidades también se situaban en la parte superior del ranking. En el caso contrario, tenemos a Extremadura que también tiene el PIB per cápita más bajo

**Tabla 1.5 Índice de producción industrial media (2008-2021)**

Situación CCAA		Media
Alta	Madrid	135,96
	Aragón	130,63
	Canarias	130,41
	País Vasco	130,28
Estándar	Andalucía	127,88
	Murcia	127,87
	Rioja	127,67
	Castilla - La Mancha	127,41
	Asturias	127,16
	Galicia	125,73
	Comunidad Valenciana	124,60
	Cataluña	124,56
Castilla y León	122,78	
Baja	Extremadura	119,62
	Navarra	119,59
	Cantabria	119,35
	Baleares	99,17
Cuartiles	Q1	121,20
	Mediana	127,16
	Q3	129,08

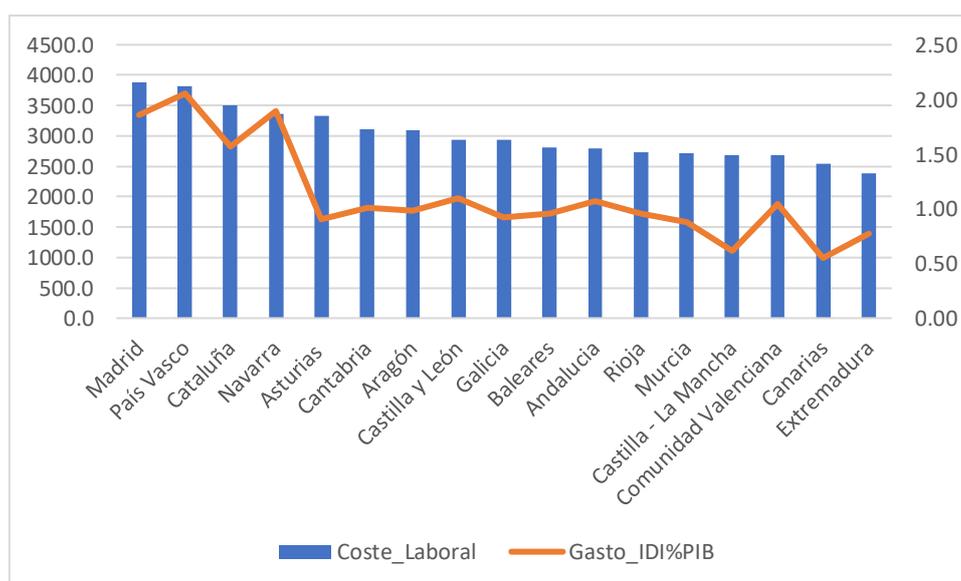
*Fuente: Elaboración propia a partir de datos del INE.*

Es sorprendente como Navarra tiene un IPI bajo en comparación con las demás

regiones a pesar de ser de las más ricas en términos de riqueza por habitante y ser la que menor riesgo de pobreza tiene. Es posible que la economía de Navarra no dependa tanto de la producción industrial como de otros sectores (servicios educativos o sanitarios), que pueden ser motores importantes del crecimiento económico de la región. Por lo tanto, una baja producción industrial no necesariamente significa una economía débil o poco desarrollada.

A continuación, analizamos, en el gráfico 1.13, el coste laboral y el gasto en investigación en porcentaje del PIB con el objetivo de estudiar si existe relación entre estas dos variables, relacionadas con el desarrollo económico. Por un lado, el coste laboral es una parte muy relevante de la estructura de costes de una empresa y de su competitividad y por otro lado, el gasto en investigación es un indicador de la capacidad tecnológica, que puede también afectar a la competitividad y mejorar la capacidad para generar riqueza y empleo. Las Comunidades están ordenadas de mayor a menor coste laboral. Podemos observar que las Comunidades que tienen mayor coste laboral también tienen mayor gasto en investigación y mayor IPI, como Madrid y País Vasco. Parece que la innovación contribuye a aumentar la producción industrial.

**Gráfico 1.13 Coste Laboral y Gasto en investigación en porcentaje del PIB por CCAA (2008-2021)**



*Fuente: Elaboración propia a partir de datos del INE.*

Además, las Comunidades Autónomas con mayores costes, son las mismas que

tienen mayor gasto en investigación y las que tienen el IPI más alto como Madrid y País Vasco. Parece que la innovación contribuye a aumentar la producción industrial.

Una vez finalizado el estudio descriptivo de algunas de las principales variables sociodemográficas y económicas, podemos señalar qué horizonte temporal resulta adecuado para analizar estadísticamente.

En primer lugar, el año 2010 es interesante ya que está marcado por la crisis del 2008 y es donde se comienzan a ver las diferencias en Población y envejecimiento. El año 2015 es el momento en el que podemos decir que la economía entra de nuevo en fase expansiva, con nuevas tendencias en las variables y divergencias entre regiones después de la crisis. Este será el segundo año de estudio. Finalmente, situaré el último análisis en 2021, ya que es el último año del que se tienen datos definitivos y podremos analizar las consecuencias de la pandemia por Comunidades Autónomas y cómo esto ha influido en la sociedad<sup>9</sup>.

---

<sup>9</sup> Aunque el intervalo que he elegido es de cinco años y la secuencia lógica tras el 2010 y 2015, sería el año 2020, he decidido obviar este año y “saltar” a 2021 por ser aquel un año muy especial, marcado por la irrupción del Covid-19, con el shock y la alteración que ello supuso. Elegir 2020 era arriesgar conclusiones e interpretaciones.

## **2. ANÁLISIS FACTORIAL Y ANÁLISIS DE CLÚSTER**

El objetivo del análisis descriptivo del apartado anterior fue identificar las variables relevantes que permiten contextualizar la situación española y determinar las posibles diferencias entre Comunidades Autónomas. Además, se han llevado a cabo estudios de las posibles relaciones entre las variables, con el propósito de obtener conclusiones más precisas.

En la segunda parte de este trabajo, se explicarán teóricamente los análisis estadísticos que se aplicarán posteriormente. En primer lugar, se explica qué y cómo se identifican los factores que resumen las variables, para después utilizar las puntuaciones factoriales obtenidas en un análisis clúster o de conglomerados, que tiene como objetivo agrupar las Comunidades Autónomas según las semejanzas de sus características y separar aquellas que presentan mayores diferencias.

### **2.1 ¿Qué es un Análisis Factorial?**

El análisis factorial es una técnica de reducción de datos multivariante utilizada para explicar correlaciones entre variables observadas. Después se crean combinaciones lineales para agrupar estas variables en un conjunto de datos menor al original con variables no observadas llamadas factores o componentes.<sup>10</sup>

El objetivo de realizar un análisis factorial es la reducción de un conjunto de datos original a otro más pequeño buscando el mínimo número de factores o componentes, perdiendo la mínima información posible de las variables explicadas en la varianza. Esta reducción de datos se realiza en función de la correlación lineal de las variables basadas en la diagonalización de la matriz de covarianzas o correlaciones. Estos factores creados y sus puntuaciones factoriales serán utilizados para la realización del análisis clúster y para las interpretaciones de los conglomerados.

El análisis factorial puede ser de dos tipos: el exploratorio que consiste en determinar el número de factores del análisis ya que no se conocen y en caso contrario, el confirmatorio que se realizará si ya conocemos los factores y para

---

<sup>10</sup> Durante la elaboración de este trabajo, me referiré indistintamente a factores o componentes, significando lo mismo.

comprobar su veracidad realizamos hipótesis.<sup>11</sup>

En este trabajo se utilizará el análisis factorial exploratorio, dado que no se cuenta con información previa acerca del número de factores necesarios para resumir la información de las variables. Será el programa SPSS, quien determine el número adecuado de los mismos para después realizar el análisis clúster.

Para la realización del análisis factorial se deben seguir los siguientes pasos:

En primer lugar, se debe determinar la pertinencia del análisis factorial. Para ello se utilizan la prueba de Esfericidad de Bartlett y el índice de Kaiser-Meyer-Olkin.

La prueba de Esfericidad de Bartlett es utilizada para comprobar si las varianzas de las muestras si son iguales de tal forma que se presenta en esta hipótesis:

H0: Matriz de correlaciones es igual a la matriz identidad

H1: Alguno de los coeficientes de correlación es distinto de cero<sup>12</sup>

En definitiva, buscamos que el p-valor sea  $< 0.05$  para rechazar la H0 y aceptar la H1, que explicaría que las variables tienen alta correlación y se puede continuar con el análisis factorial.

El índice de Kaiser-Meyer-Olkin, compara los coeficientes de correlación lineal con las parciales de tal forma que esta medida varía entre cero y uno. Por tanto, si hay presencia de factores comunes donde el análisis factorial es adecuado implica que la correlación parcial es pequeña en comparación con la lineal, entonces el KMO estaría cerca de uno. Los valores críticos del índice de Kaiser-Meyer-Olkin son los siguientes:

- ✓  $KMO < 0,5 \Rightarrow$  Inaceptable ya que implicaría que no hay presencia de factores comunes indicando que existe poca correlación entre el conjunto de datos que estamos analizando
- ✓  $KMO \geq 0,5 \Rightarrow$  Aceptable por lo tanto podemos realizar el análisis sin problema.
- ✓  $KMO \geq 0,75 \Rightarrow$  Perfecto, además a medida que este valor tienda a la unidad es mejor, indicando correlación y presencia de factores

---

<sup>11</sup> De la fuente Fernández, S. (2011), "Análisis factorial".

<sup>12</sup> Manríquez Gary, J.C (2022), "Técnicas Estadísticas multivariantes para la investigación en la administración" (pp. 16-17)

comunes<sup>13</sup>.

Si ambas pruebas (KMO y Bartlett) son aptas para el análisis, esto implica que existe alta correlación entre la base de datos usada y la presencia de factores comunes, por lo tanto, la realización del Análisis factorial sería correcta.

En segundo lugar, se deben determinar el número de factores.

En este caso como vamos a realizar un análisis factorial exploratorio, necesitamos concretar el número de factores apropiado, para ello voy a seguir dos reglas o criterios para la obtención de estos:

- a) La Regla de Kaiser en la cual conservamos aquellos autovalores de la matriz de correlación que sean superiores a la unidad y descartamos aquellos factores o componentes que quedan por debajo, asumiendo esa pérdida de información.
- b) El Criterio de porcentaje de la varianza donde buscamos que la acumulación de la información explicada por esos factores (varianza acumulada) sea superior al 70%.<sup>14</sup>

Es decir, bajo estos dos criterios seleccionamos el número de componentes o factores óptimo siempre que estos componentes tengan un autovalor superior a uno y que el conjunto de los componentes seleccionados sume al menos el 70% de la varianza acumulada, todos los demás quedarán descartados y perderemos esa información del conjunto de datos.

Después, debemos analizar la **tabla de Comunalidades**, la cual muestra todas las variables analizadas en puntuaciones que van de cero a uno donde explican la varianza de los factores comunes. Esto significa que las variables que estén mejor representadas en los factores tenderán a la unidad mientras que si las variables tienen una carga inferior de  $|0,6|$  estarán “mal” representadas y puede originar problemas en el análisis por lo que es conveniente eliminar la variable<sup>15</sup>

A continuación, se debe interpretar la **matriz de componentes** la cual nos concede la solución factorial. Usaré la rotación varimax que lo explicaré en el

---

<sup>13</sup> De la fuente Fernández, S. (2011), “Análisis factorial”.

<sup>14</sup> Mercedes Prieto, M. (2021), “Análisis Factorial”

<sup>15</sup> De la fuente Fernández, S. (2011), “Análisis factorial”.

punto posterior. Respecto a la interpretación de la solución factorial seleccionaré la **matriz de componentes rotados**.

Dentro de esta tabla aparecerán cada variable con puntuaciones factoriales para cada componente las cuales pueden ser negativas o positivas y oscilan entre menos uno y uno. Los valores negativos implican que cuanto mayor sea esa variable menor puntuación recibirá ese componente, en caso contrario si la variable tiene una puntuación factorial positiva indicaría que a mayor valor de esa variable más aumentará ese componente.

Las puntuaciones solo se interpretan si están por encima de  $|0,6|$  y cuánto más cerca estén de la unidad implica que afecta más a ese componente. Además, cada variable solo formará parte del componente con mayor puntuación obtenida para facilitar la interpretación a pesar de que afecte a todos.

Por último, tenemos las puntuaciones factoriales de cada variable asociada a un componente, es el momento en el que tenemos que dar sentido a cada factor agrupado por cada variable por lo que en este momento los componentes adoptarán un nombre que de sentido al factor.

Finalmente, para la realización del análisis factorial usaré el Método de Componentes Principales con la Rotación varimax y Puntuaciones Factoriales Anderson-Rubin.

El **Método de Componentes Principales** consiste en formar combinaciones lineales para obtener la solución factorial inicial. El primer componente será el de mayor varianza y a medida que tengamos más componentes esa varianza irá disminuyendo, además estos componentes no están correlacionados.

La **Rotación Varimax con Puntuaciones Factoriales de Anderson-Rubin** tiene como objetivo mejorar la interpretación de las variables en los factores. Para mejorar la interpretación se selecciona una variable minimizando las altas cargas factoriales situándolas cerca de la unidad para un solo factor, mientras que sitúa bajas cargas factoriales para el resto de los factores, así, esa variable será fácilmente interpretada al estar en un solo factor con una puntuación muy alta mientras que en los demás factores está interpretada con una baja puntuación. En definitiva, lo que se busca con esta rotación y las puntuaciones factoriales es maximizar una variable de tal forma que quede representada en

un componente y descartada en todos los demás, así mejoramos la interpretación de los factores.<sup>16</sup>

## 2.2 Análisis Clúster

El Análisis Clúster es una técnica multivariante de agrupación de casos en conglomerados o clúster, siendo el objetivo de este análisis que los grupos creados se agrupen en variables que les caractericen por ser lo más parecidos posibles (homogéneos) y la mayor diferencia posible entre los grupos (heterogéneos).

Esta técnica estadística nos permite la creación de grupos en función del conjunto de datos, es decir a priori nosotros no conocemos los grupos y tenemos que considerar a partir del análisis cual es la solución más óptima para nuestro ejercicio. Para la realización del análisis usaré el método jerárquico con la agrupación en clústeres por enlace entre grupos en medida de distancia euclídea al cuadrado que explicaré a continuación.<sup>17</sup>

Para la realización del análisis clúster es necesario realizar con anterioridad el análisis factorial exploratorio, el cual nos dará nuevas variables que es la solución factorial y nos va a servir para que los conglomerados los agrupe en función de estas puntuaciones. Por otra parte, los casos los etiquetamos por la variable CCAA ya que queremos que nos agrupe aquellas Comunidades Autónomas que se parecen entre ellas separándolas de las que tienen diferencias significativas.

Los métodos que podemos usar para la realización del análisis clúster destacan el método jerárquico y el no jerárquico o de K-medias.

Respecto al **Método Jerárquico** consiste en unir conglomerados próximos entre sí para formar uno más grande. El proceso por el cual se unen está dispuesto en etapas y queda registrado en un gráfico llamado dendrograma<sup>18</sup>.

La medida de distancia que usaré será la euclídea al cuadrado, para ello, en la matriz de distancias realizamos la resta al cuadrado y la sumamos. En función de estas medidas se juntarán y se separarán los casos en clústeres dado su

---

<sup>16</sup> Montanero Fernández, J. (2008), Capítulo 11: "Análisis factorial, Análisis multivariante"

<sup>17</sup> Peña, D. (2002), Capítulo 8: "Análisis de conglomerados, Análisis de datos multivariantes"

<sup>18</sup> Montanero Fernández, J. (2008), Capítulo 12: "Análisis clúster, Análisis multivariante"

similitud o diferencias

Método de enlace entre grupos, donde la distancia entre dos grupos es la media de la distancia entre los posibles individuos de cada grupo. Existen otros métodos de enlaces más utilizados como el método de Ward que se encarga de minimizar las varianzas totales de cada clúster, pero tiende a crear conglomerados de tamaño similar lo cual no es interesante para este caso ya que queremos que si una Comunidad Autónoma es muy diferente a las demás nos la separe y no nos la clasifique “mal”.

El **Método de K-Medias es No Jerárquico** donde el número final de conglomerado está determinado y se aglomeran entorno a semillas, las cuales determinan la proximidad en función de los casos que más se asemejan.

Este método se realiza con los centros de los grupos iniciales seleccionados a priori donde se calculan las distancias euclídeas de cada grupo todas las veces que se introduzca un nuevo elemento (nueva media de grupo), finalmente se pone fin al proceso si no se consiguen soluciones inferiores a los puntos que ya hemos tenido en cuenta.<sup>19</sup>

Finalmente, usaré el análisis jerárquico únicamente ya que para la realización del K-medias primero tenemos que hacer el jerárquico y vamos a obtener conclusiones similares por lo que el trabajo se haría repetitivo.

A continuación, vamos a seleccionar el número de conglomerados y para ello tenemos que establecer una buena selección de clúster mediante el Gráfico **Dendrograma** (solo está disponible para el método jerárquico), donde nos solemos fijar en la etapa comprendida entre la cinco y la diez donde observamos que los conglomerados tienden a tamaños homogéneos.

Por otra parte, usamos otras dos medidas (sirven para ambos métodos) como son el **Índice de Calinski-Harabasz (CH)** y el **Índice de Silueta**. Este primero se obtiene con la división entre la suma de cuadrados entre-grupos (explicada) y la suma de cuadrados intra-grupos (residual), el SPSS nos la facilita con medias cuadráticas y realizamos el cálculo sobre un Excel para cada solución de Clúster que hemos determinado. El valor más alto implica que es mejor esa

---

<sup>19</sup> Peña, D. (2002), Capítulo 8: “Análisis de conglomerados, Análisis de datos multivariantes”

solución de número de clústeres.

Por otro lado, el Índice de silueta se calcula desde el propio SPSS el cual nos facilita una media total donde también cogemos el valor más alto, además este índice nos facilita otras interpretaciones ya que oscila entre menos uno y uno, siendo más próximo a la unidad como bien clasificado ese individuo en su clúster y mal clasificado cuando este cerca de menos uno lo que quiere decir que ese individuo está más cerca de otro clúster que del clasificado.<sup>20</sup>

En las tablas del ejercicio práctico el índice de silueta vendrá representado por “**S6**”: significa valor de la silueta y el número será la solución de clústeres que en este caso sería seis. Lo que se busca con esta interpretación es que los valores de cada Comunidad Autónoma este lo más cerca posible de uno ya que estaría “bien clasificado” y en caso contrario si está cerca de menos uno estaría “mal clasificado”. En la tabla también aparecerá “**SM6**” lo que significa siguiente mejor clúster y el número será la solución de clústeres que en este caso también será seis. En este caso lo que interpretamos es que, si el clúster al que la Comunidad Autónoma que estamos observando no existiera, ¿A que otro clúster pertenecería? Es decir, nos asigna el siguiente clúster al que estaría mejor clasificado, así podemos ver que clústeres y que Comunidades Autónomas se parecen más o menos.

Al final de cada caso se realizará un **Resumen de Caso** con el objetivo de obtener interpretaciones y conclusiones. Este resumen de caso estará compuesto por los clústeres que se han formado a través de las puntuaciones factoriales en media para comprobar cómo afectan a cada clúster. Es decir, esta tabla nos va a facilitar interpretar como los individuos se agrupan en distintos conglomerados y que características tienen estos según los componentes que hemos formado. Los valores vendrán representados con números negativos o positivos. Los negativos indican que ese componente o factor al clúster que estamos observando afecta de forma negativa, es decir que los valores en media son más bajos que en los demás. Por el contrario, los valores serán más altos de lo normal si aparecen de forma positiva. Es decir que podemos observar también cual es el mejor y peor clúster asociado a un componente en concreto.

---

<sup>20</sup> Mercedes Prieto, M. (2021), “Análisis de conglomerados”.

### **3. ANÁLISIS MULTIVARIANTE APLICADO A LAS CCAA.**

A continuación, voy a realizar los análisis multivariantes para intentar obtener algunas conclusiones relativas a similitudes y diferencias entre las Comunidades Autónomas españolas.

En primer lugar, tenemos tres casos distintos, correspondientes a los tres años de estudio 2010, 2015, 2021 en los cuales vamos a seguir la misma metodología y los mismos pasos. Para cada uno de los años, voy a realizar un Análisis factorial para resumir las 14 variables en componentes o factores separados por una muestra de 17 Comunidades Autónomas.

Una vez obtenidos los factores, guardamos estas soluciones factoriales que se convertirán en “nuevas variables” que utilizaré para realizar un Análisis Clúster Jerárquico etiquetando por casos en la variable CCAA. Esto nos permitirá agrupar las Comunidades Autónomas que tengan similitudes y separar aquellas que tengan mayores diferencias en función de las puntuaciones factoriales.

#### **3.1 Elección y tratamiento de las variables**

En la selección de las variables el criterio ha sido el análisis descriptivo realizado con anterioridad. Este análisis permite que busquemos las variables que están correlacionadas para conseguir un conjunto de datos que nos permita realizar el análisis factorial y clúster.

Las variables utilizadas han sido recogidas del INE (2022), después han sido introducidas en IBM SPSS Statistics 26 (2022) donde se realizan los análisis factoriales y clústeres, respectivamente.

Como ya señalamos, las variables recogen la información de las Comunidades Autónomas españolas con excepción de Ceuta y Melilla debido a la escasa información que existe de ellas. Los años en los cuales vamos a realizar el estudio serán 2021, 2015 y 2010, considerando que en estos años establecemos un horizonte temporal donde podremos observar las diferencias significativas que hemos encontrado en el análisis descriptivo.

Las variables escogidas se pueden agrupar en dos grandes grupos:

1. Variables Macroeconómicas: IPC, IPV, IPI, PIB, Rentas y gastos medios, tasa de pobreza y tasa de paro y por último el coste laboral y

el gasto en investigación.

El PIB per cápita, que fue incluido en el análisis descriptivo, ha sido eliminada en los análisis multivariantes debido a su excesiva correlación con el PIB y la Población, empeorando los análisis.

2. Variables socioculturales y demográficas: este tipo de variables nos va a proporcionar información de la sociedad y la población. Son el saldo vegetativo, población y edad media de la población.

Algunas de las variables están correlacionadas entre sí<sup>21</sup>, como se puede observar en las Tablas A.6, A.7, A.8 de anexos. Estas tablas han sido obtenidas gracias a la correlación Parcial de Pearson que nos indica, de menos uno a uno, si existe correlación negativa o positiva entre dos variables. Los valores cercanos a la unidad indican correlación, con un valor negativo, la relación es inversamente proporcional y proporcional en el caso contrario.

Una vez analizadas todas las variables por la Correlación de Pearson, procederíamos a realizar el análisis factorial y tendríamos que tener en cuenta la pertinencia del análisis como el KMO y la esfericidad de Bartlett.

En la tabla 3.1 se presentan las 14 variables seleccionadas, con su correspondiente abreviatura utilizada en el conjunto de datos de SPSS. Además, se proporciona información sobre la fuente de los datos y la observación correspondiente. Cabe destacar que en el apartado de observación se hacen referencias a los gráficos utilizados en el análisis descriptivo de la situación de las CCAA, con el fin de resumir y organizar todas las manipulaciones realizadas en dicha sección.

---

<sup>21</sup> Algunas variables que veremos en los casos prácticos pueden ser descartadas debido a su no correlación con las demás.

**Tabla 3.1 Variables**

Variable	Abreviatura	Observación	Fuente
Saldo vegetativo en porcentaje	S_Vegetativo	Crecimiento vegetativo: natalidad y mortalidad 2007-2020	INE
Índice de precios al consumo	IPC_Índice	Tendencia del IPC desde 2008-2021	INE
Índice de precios industriales	IPI_Índice	Tasa de variación anual del IPC 2008-2021	INE
Índice de precios de vivienda	IPV_Índice	IPI en media. 2008-2021	INE
		Tendencia del IPV desde 2008-2021	INE
		Tasa de variación anual del IPV 2008-2021	INE
		PIB Pm en media 2002-2021	INE
PIB a precios de mercado	PIBpm	Tasa de variación anual del PIB 2002-2021	INE
		Tasa de variación anual del PIB por CCAA 2002-2021	INE
		PIB per cápita en media 2002-2021	INE
		Población en media 2002-2021	INE
		Tasa de Variación anual de la población por CCAA 2002-2021	INE
Población	Población	Tasa de Variación anual de la población y edad media de la población 2002-2021	INE
		Tasa de variación anual de las CCAA que pierden población 2002-2021	INE
		Tendencia de rentas y gastos mediosde la CCAA 2008-2021	INE
Renta media del hogar	Renta_Media_H	Tasa de variación anual de rentas y gastos medios 2008-2021	INE
		Tendencia de rentas y gastos mediosde la CCAA 2008-2021	INE
Renta media por unidad de consumo	Renta_M_UnidC	Tasa de variación anual de rentas y gastos medios 2008-2021	INE
		Tendencia de la Tasa de Paro y riesgo de pobreza 2008-2021	INE
Tasa de paro	T_Paro	Tasa de Variación anual de la tasa de paro y riesgo de pobreza por CCAA 2008-2021	INE
		Tendencia de la Tasa de Paro y riesgo de pobreza 2008-2021	INE
Riesgo de pobreza	T_R_Pobreza	Tasa de Variación anual de la tasa de paro y riesgo de pobreza por CCAA 2008-2021	INE
		Edad media de la población en media 2002-2021	INE
Población edad media	Edad_Media	Tasa de Variación anual de la población y edad media 2002-2021	INE
		Tasa de variación de la edad media de la población por CCAA 2002-2021	INE
		Tendencia de rentas y gastos mediosde la CCAA 2008-2021	INE
Gasto Medio por persona	Gasto_M_Persona	Tasa de variación anual de rentas y gastos medios 2008-2021	INE
Gasto en investigación y desarrollo en porcentaje del PIB	G_I+d_PIB	Tendencia del coste laboral y gasto en investigación en porcentaje del PIB por CCAA 2008-2021	INE
Coste laboral industrial	C_Laboral_ind	Tendencia del coste laboral y gasto en investigación en porcentaje del PIB por CCAA 2008-2021	INE

*Fuente: Elaboración propia*

Estas variables vienen etiquetadas por tipo “escala” lo que implica que tienen unos valores numéricos. Además, hay tres conjuntos de datos distintos, 2010, 2015 y 2021, agrupados todos ellos por una variable de tipo nominal, “CCAA” donde están todas las Comunidades Autónomas españolas, excepto Ceuta y Melilla.

La variable CCAA representa las Comunidades Autónomas de España, con la excepción de Ceuta y Melilla. Esta variable se combina con las otras 14 variables del conjunto de datos (variables macroeconómicas y sociodemográficas), lo que nos permite etiquetar los casos en el análisis clúster y, de este modo, segmentar los diferentes grupos que se generan. Es decir, cada Comunidad está compuesta por 14 variables nombradas en la tabla 3.1

Se pueden consultar los conjuntos de datos finales obtenidos y las nuevas variables creadas que he utilizado para realizar el estudio multivariante en el Anexo (Tablas A.9, A.10, A.11 correspondientes a 2010, 2015 y 2021, respectivamente).

### 3.2 Caso 1: año 2010

En este primer caso debemos tener en cuenta los acontecimientos revelados en el análisis descriptivo que puede consultarse en: (1.2 “Análisis descriptivo”).

A la vista de la tabla 3.2, observamos que el KMO es superior 0,5 y en la Prueba de esfericidad de Bartlett el p-valor tiende a cero por lo tanto podemos concluir que el análisis factorial con este conjunto de datos es adecuado ya que hay presencia de factores comunes.

**Tabla 3.2 Pertinencia Del Análisis**

Medida Kaiser-Meyer-Olkin	KMO	0,564
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	303,587
	gl	91
	Sig.	0

*Fuente: Elaboración propia*

**Tabla 3.3 Varianza Total Explicada**

Componente	Suma de cargas al cuadrado de la extracción			Suma de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	6,744	48,169	48,169	6,611	47,219	47,219
2	3,46	24,716	72,885	2,55	18,213	65,432
3	1,298	9,271	82,156	2,341	16,724	82,156

*Fuente: Elaboración propia*

La varianza total explicada nos permite resumir las 14 variables en la solución óptima de tres componentes o factores.

Cada factor explica el 47,2%, 18,2% y 16,7%<sup>22</sup> de la varianza total explicada, consiguiendo en el porcentaje acumulado un 82%, lo que implica que se ha perdido poca información por agrupar variables en factores.

<sup>22</sup> A mayor porcentaje de varianza total implica que ese componente tiene más peso en el análisis que los de menor varianza. Por tanto, siempre nuestro primer componente será el más importante.

**Tabla 3.4 Comunalidades**

Variables	Extracción
S_Vegetativo	0,955
IPC_Índice	0,412
IPI_Índice	0,586
IPV_Índice	0,575
PIBpm	0,849
Población	0,796
Renta_Media_H	0,954
Renta_M_UnidC	0,953
T_Paro	0,911
Riesgo_pobreza	0,913
Edad_Media	0,984
Gasto_M_Persona	0,916
G_I+D_PIB	0,762
C_Laboral_ind	0,936

*Fuente: Elaboración propia*

En la Tabla de Comunalidades podemos observar, en rojo, aquellas variables que peor quedan recogidas y que son; “IPC”, “IPI”, “IPV”. Ciertamente, podrían ser eliminadas de este análisis sin ningún problema, pero las mantendré porque mejorarán las interpretaciones y conclusiones. Por el contrario, de color verde tenemos las de mejor representación “S\_Vegetativo”, “Edad\_Media”.

**Tabla 3.5 Matriz De Componentes Rotados**

Variables	Componente		
	1	2	3
S_Vegetativo			0,957
IPC_Índice	-0,624		
IPI_Índice		0,712	
IPV_Índice		-0,651	
PIBpm		0,772	
Población		0,826	
Renta_Media_H	0,969		
Renta_M_UnidC	0,965		
T_Paro	-0,774		
Riesgo_pobreza	-0,89		
Edad_Media			-0,942
Gasto_M_Persona	0,947		
G_I+D_PIB	0,869		
C_Laboral_ind	0,939		

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

a La rotación ha convergido en 5 iteraciones.

*Fuente: Elaboración propia*

En la matriz de componentes rotados han sido eliminadas aquellas variables con un coeficiente de valor absoluto inferior a 0.6.

El primer factor recoge variables relacionadas con la situación económica de las CCAA donde destacan las rentas y gastos medios que afectan positivamente, lo que podría indicar “Rentas Altas”. Además, también se incluyen otras variables macroeconómicas como la tasa de paro o riesgo de pobreza que afectan negativamente, es decir, a mayor tasa de paro o riesgo de pobreza menor valor de este componente. Por tanto, podemos concluir que este componente nos indica la **“Buena situación Económica”**.

El segundo factor recoge variables relacionadas con la población y el PIB. A pesar de que también se incluye el IPV, la descarto por su poco peso, como hemos observado en la tabla de comunalidades. Por tanto, podemos decir que este factor recoge atributos relacionados con la **“Industrialización y elevada población”**. En el Análisis descriptivo ya se observó la relación directa entre población y PIB, a más población, más PIB.

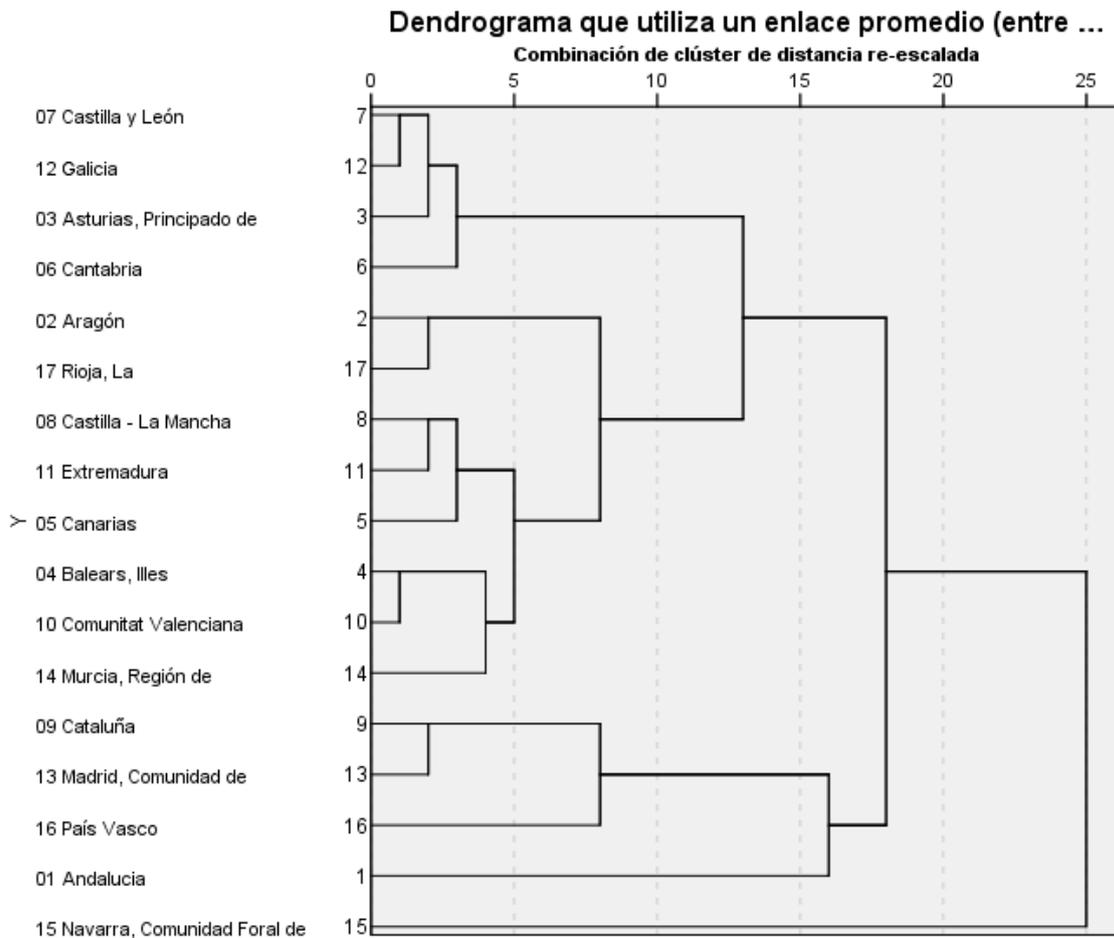
Por último, el tercer factor recoge tan solo dos variables, saldo vegetativo y edad media de la población de forma negativa, esto implica que este factor recoge aquellas comunidades que tienen un saldo vegetativo favorable y una edad media de su población baja, por tanto, este factor será de **“Población no envejecida”**.

Con estas puntuaciones factoriales, vamos a realizar el Análisis Clúster Jerárquico, con la intención de mejorar la interpretación de los datos, las puntuaciones factoriales han sido renombradas de la siguiente manera:

Factor 1: **“Buena situación Económica”**.

Factor 2: **“Industrialización y elevada población”**.

Factor 3: **“Población no envejecida”**.



*Fuente: Elaboración propia*

El dendrograma facilita el rango de soluciones clúster que debemos escoger, como observamos entre la etapa 5 y la 10, podríamos seleccionar entre 3 y 6 clústeres.

Vamos a buscar cuál es la mejor solución para la interpretación de los datos a través del Índice de Silueta y el CH.

**Tabla 3.6 Determinación Del Número De Clústeres**

Nº Clúster	3	4	5	6
Índice Silueta	0,41	0,43	0,55	0,61
CH	5,2	4,7	9,3	11,3

*Fuente: Elaboración propia*

Elegimos el número de Clúster con el índice de silueta y el CH más alto, por tanto, la mejor solución para este caso son seis Clústeres.

**Tabla 3.7 Caso De Clúster**

<b>Caso</b>	<b>6 clústeres</b>
Andalucía	1
Aragón	2
Asturias, Principado de	3
Baleares, Illes	4
Canarias	4
Cantabria	3
Castilla y León	3
Castilla - La Mancha	4
Cataluña	5
Comunitat Valenciana	4
Extremadura	4
Galicia	3
Madrid, Comunidad de	5
Murcia, Región de	4
Navarra, Comunidad Foral	6
País Vasco	5
Rioja, La	2

*Fuente: Elaboración propia*

En la tabla 3.7 podemos observar la agrupación de Comunidades Autónomas en conglomerados, en este caso destacamos que hay dos Comunidades que quedan agrupadas como independientes que son Andalucía (1) y Navarra (6), esto implica que estas dos CCAA, en el 2010 tuvieron muchas discrepancias con las demás de tal forma que ninguna de las dos se parece a las comunidades restantes.

Estas discrepancias pueden verse observadas en la tabla 3.8, donde Andalucía tiene una mala situación económica, pero con elevada población e industrialización. El otro conglomerado con mala situación económica, como Andalucía, es el clúster 4 donde sus características como industrialización y población son diferentes ya que estos atributos son estándares.

En Navarra ocurre exactamente lo mismo, pero al contrario. Es decir, su situación económica es muy favorable pero su población e industrialización es muy baja y el conglomerado (5) que tiene la mejor situación económica coincide con lo contrario, que tiene mucha industrialización y elevada población.

En ambos casos, estas diferencias significativas permiten separar a ambas

comunidades en un conglomerado individual para Navarra y otro para Andalucía. Con otro conjunto de datos, la solución propuesta podría no ser apropiada debido a la diferencia en el tamaño muestral entre los clústeres. En algunos clústeres, puede haber una sola comunidad como el de Andalucía y Navarra, mientras que, en otros, la muestra es mucho más grande como en el clúster 4. Sin embargo, en este caso particular, la solución adoptada nos permite concluir que hay una gran disparidad entre las Comunidades Autónomas, a pesar de ser parte del mismo país.

**Tabla 3.8 Resumen De Casos**

<b>Cluster</b>		<b>Buena Situación Económica</b>	<b>Industrialización y Elevada Población</b>	<b>Población No Envejecida</b>
<b>1</b>	Media	-0,63	2,04	0,55
<b>2</b>	Media	-0,01	-1,23	-0,13
<b>3</b>	Media	-0,05	0,13	-1,47
<b>4</b>	Media	-0,87	-0,10	0,62
<b>5</b>	Media	1,52	0,85	0,32
<b>6</b>	Media	1,49	-2,05	0,92

*Fuente: Elaboración propia*

El resumen de informe de casos nos permite comparar en media como cada clúster, formado por las puntuaciones factoriales, varía en función de esas características propias.

El clúster 1 está formado únicamente por Andalucía, la cual nos indica que tiene una mala situación económica a pesar de estar industrializada. Su población es elevada y no está envejecida.

Andalucía es el conglomerado con mayor población e industrialización, sin embargo, como vimos en análisis descriptivo, (tabla 1.4), podemos observar cómo, a pesar de que es de las que más población y PIB tiene, su PIB per cápita es de los más bajos. Además, en el 2010 tenía el IPI de los más altos con 112.

El clúster 2 está compuesto por Aragón y La Rioja. Tienen una situación económica normal con baja industrialización. Destacan por su baja población, con una edad media de población estándar.

El clúster 3 está constituido por Asturias, Cantabria, Castilla y León y Galicia. Destacan por ser las CCAA con la población más envejecida con bastante diferencia de los demás clústeres. Respecto a la situación económica es estándar y su industrialización y número de población también. Tenemos que recordar que Asturias, Galicia y Castilla y León, comienzan en 2010 a perder población. Los resultados efectivamente arrojan que tienen la población más envejecida.

El clúster 4 es el más numeroso y está formado por Canarias, Baleares, Valencia, Murcia, Castilla La Mancha y Extremadura. Las cuales destacan por tener la peor situación económica y por su población joven. Su industrialización y número de población es normal.

El clúster 5 está formado por Cataluña, País Vasco y Madrid, siendo el único conglomerado con todos los factores positivos lo cual nos indica que son las mejores Comunidades Autónomas en prosperidad. Son las más ricas, en términos económicos, tienen elevada industrialización y población, además de no estar envejecidas.

El clúster 6 incluye únicamente a Navarra, la cual destaca por su muy favorable situación económica y su población joven, al mismo tiempo que tiene una baja industrialización y muy poca población. En este caso en concreto, como está formado por un solo individuo, podemos ir a los datos y comparar con los resultados obtenidos: Navarra tiene poca población al ser una comunidad pequeña y está muy poco industrializada. Sin embargo, observamos una contradicción ya que Navarra está situada en la mediana de la edad media de la población, a pesar de que en este análisis la sitúa como el conglomerado con la población más joven, es por tanto una Comunidad a tener en cuenta en los siguientes casos y observar su variación/comportamiento.

Como tenemos dos clústeres con una única región, vamos a utilizar el índice de silueta para comprobar si están bien clasificados y a que clúster se parecen más, es decir, vamos a ver a que clúster podrían pertenecer:

**Tabla 3.9 Índice de Silueta**

<b>CCAA</b>	<b>Clúster</b>	<b>SM6</b>	<b>S6</b>
<b>Andalucía</b>	1	4	1
<b>Aragón</b>	2	3	0,72
<b>Asturias, Principado de</b>	3	2	0,632
<b>Balears, Illes</b>	4	2	0,427
<b>Canarias</b>	4	1	0,497
<b>Cantabria</b>	3	2	0,435
<b>Castilla y León</b>	3	2	0,773
<b>Castilla - La Mancha</b>	4	2	0,457
<b>Cataluña</b>	5	1	0,675
<b>Comunitat Valenciana</b>	4	2	0,477
<b>Extremadura</b>	4	2	0,334
<b>Galicia</b>	3	2	0,769
<b>Madrid, Comunidad de</b>	5	1	0,607
<b>Murcia, Región de</b>	4	2	0,445
<b>Navarra, Comunidad Foral</b>	6	2	1
<b>País Vasco</b>	5	2	0,429
<b>Rioja, La</b>	2	4	0,751

*Fuente: Elaboración propia*

La explicación de esta tabla, así como S6 y SM6 está explicado con anterioridad en: (2.2. “Análisis clúster”).

El “S6 “nos indica, de menos uno a uno, si el clúster está bien clasificado, como no hay ninguno cerca de menos uno, esto significa que todas las CCAA están bien clasificadas.

En segundo lugar, “SM6”, nos indica cual es el siguiente mejor clúster en el que podría estar incluida cada CCAA. En este caso nos vamos a centrar en Andalucía y en Navarra, esos clústeres serían el 4 y el 2, respectivamente.

Es decir que, si Andalucía perteneciera al clúster 4 ya que es su siguiente mejor clúster, lo clasificaría como mala situación económica, industrialización y número de población estándar donde su población en edad joven es alta. Esto no sería cierto ya que Andalucía está muy poblada y tiene una industrialización muy alta

solo superada por Madrid, por tanto, entendemos que separarlo en un solo clúster es correcto.

En el caso de Navarra, si perteneciera al clúster 2, lo clasificaría como situación económica estándar con baja industria y población, y la población edad media sería estándar. Respecto al clúster 2, su situación económica se podría considerar que es errónea, ya que es bastante favorable. No obstante, en cuanto a la edad media de la población, su clasificación podría estar acertada. A pesar de tener un alto porcentaje de "Población no envejecida", según el análisis descriptivo y los datos de la base de datos, Navarra se encuentra dentro de la mediana en este aspecto. Es decir, la edad media de la población en Navarra es estándar y se encuentra dentro de los, valores normales.<sup>23</sup>

### 3.3 Caso 2: Año 2015

Se considera que en 2015 se pone fin a la crisis originada en 2008 y se comienza una época de prosperidad en España donde se han corregido muchos de los desequilibrios macroeconómicos.

Comenzamos el análisis factorial de 2015 analizando la pertinencia de análisis:

**Tabla 3.10 Pertinencia Del Análisis**

Medida Kaiser-Meyer-Olkin	KMO	0,509
	Aprox. Chi-cuadrado	326,839
Prueba de esfericidad de Bartlett	gl	91
	Sig.	0

*Fuente: Elaboración propia*

El análisis factorial exploratorio con este conjunto de datos perteneciente a 2015 es posible realizarlo, sin embargo, la pertinencia del análisis está en el límite aceptable ya que el KMO está muy cerca de 0,5 y el p-valor tiende a cero. Estos dos criterios nos permiten afirmar la existencia de factores comunes y continuar con el análisis.

---

<sup>23</sup> La clasificación de Navarra ha sido objeto de controversia debido a que su inclusión en uno u otro clúster ha sido cuestionada por los componentes de población no envejecida y situación económica. No obstante, el análisis de la tabla 3.3 indica que Navarra está bien clasificada, ya que se ha considerado más importante el componente de situación económica.

**Tabla 3.11 Varianza Total Explicada**

Componente	Suma de cargas al cuadrado de la extracción			Suma de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	6,149	43,921	43,921	6,096	43,543	43,543
2	4,12	29,429	73,35	3,059	21,849	65,392
3	1,738	12,412	85,762	2,852	20,37	85,762

*Fuente: Elaboración propia*

La varianza total explicada y los autovalores son superiores a la unidad, esto implica que el componente uno explica el 43,5% de la varianza, 21,8% el componente dos y el último un 20,37%. Por tanto, el conjunto de datos analizados nos permite obtener buena información ya que pierde muy poco al resumir las variables en puntuaciones factoriales.

**Tabla 3.12 Comunalidades**

Variables	Extracción
S_Vegetativo	0,91
IPC_Índice	0,529
IPI_Índice	0,765
IPV_Índice	0,66
PIBpm	0,924
Población	0,894
Renta_Media_H	0,965
Renta_M_UnidC	0,967
T_Paro	0,887
Riesgo_pobreza	0,963
Edad_Media	0,949
Gasto_M_Persona	0,937
G_I+D_PIB	0,748
C_Laboral_ind	0,908

*Fuente: Elaboración propia*

Podemos observar en esta tabla de comunalidades el porcentaje de varianza de cada variable lo que explica a los tres factores donde destaca “Renta\_M\_UnidC” y “Edad\_Media”, por el contrario, “IPV\_Índice” y “IPC\_Índice” no tienen unos resultados satisfactorios para el análisis. La variable IPC podría ser eliminada para mejorar la pertinencia del análisis, sin embargo, se mantendrá para mejorar las interpretaciones posteriores.

**Tabla 3.13 Matriz De Componentes Rotados**

Variables	Componente		
	1	2	3
<b>S_Vegetativo</b>		0,948	
<b>IPC_Índice</b>		-0,659	
<b>IPI_Índice</b>			0,865
<b>IPV_Índice</b>		0,742	
<b>PIBpm</b>			0,852
<b>Población</b>			0,879
<b>Renta_Media_H</b>	0,973		
<b>Renta_M_UnidC</b>	0,983		
<b>T_Paro</b>	-0,901		
<b>Riesgo_pobreza</b>	-0,953		
<b>Edad_Media</b>		-0,907	
<b>Gasto_M_Persona</b>	0,967		
<b>G_I+D_PIB</b>	0,733		
<b>C_Laboral_ind</b>	0,856		

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

a La rotación ha convergido en 5 iteraciones.

*Fuente: Elaboración propia*

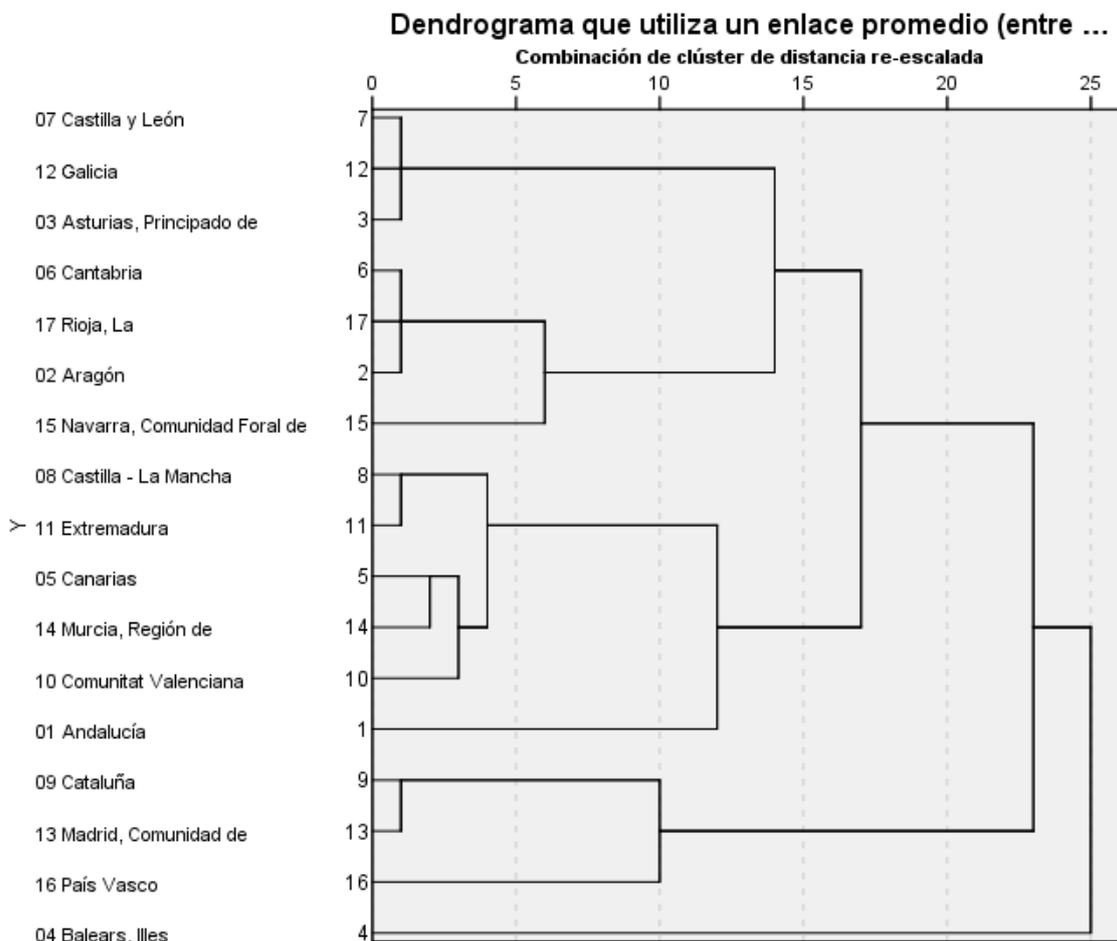
Como podemos observar las variables representadas en este primer factor son las mismas que en el anterior caso: tasa de paro, rentas y gastos medios. Por tanto, las variables resumidas nos dan información sobre la: **“Buena situación económica”**.

El segundo factor lo agrupamos como **“Población no envejecida con alto precio de las viviendas”** ya que afecta positivamente el Índice de precios de vivienda. Sin embargo, como hemos visto en la tabla de comunalidades, esta variable no tiene especial importancia al igual que el IPC, así que cuando hablemos de este factor nos referimos principalmente a población no envejecida.

El tercer factor recoge las variables que, al igual que en el caso anterior, vienen representadas por el índice industrial, la población y el PIB. Por tanto, vamos a denominarla como **“Industrialización y elevada población”**.

Finalmente, en discrepancia con el caso 1 y 2, la única diferencia que existe es que los factores han cambiado, pero siguen estando representados por las mismas variables, excepto el Índice de precios vivienda que ahora viene representado junto a la población envejecida. Además, debemos tener en cuenta

que el factor 2 en este caso tiene una varianza explicada mayor que el factor 3, entonces, tienen mejor información y es más relevante para el análisis.



*Fuente: elaboración propia*

El dendrograma nos permite buscar rango de soluciones de clúster entre cuatro y seis que podemos observar entre la etapa cinco y diez donde se forman los clústeres más homogéneos en cuanto a muestra de CCAA se refiere, aunque, finalmente se va a determinar por el Índice de silueta y el CH.

**Tabla 3.14 Determinación Del Número De Clúster**

Nº Clúster	4	5	6
Índice Silueta	0,50	0,63	0,68
CH	7,8	11,9	14,4

*Fuente: Elaboración propia*

Dado que los valores más altos de estos dos criterios corresponden a los seis clústeres será la solución única para analizar este caso.

**Tabla 3.15 Caso De Clúster**

<b>Caso</b>	<b>6 clústeres</b>
Andalucía	1
Aragón	2
Asturias, Principado de	3
Baleares, Illes	4
Canarias	5
Cantabria	2
Castilla y León	3
Castilla - La Mancha	5
Cataluña	6
Comunitat Valenciana	5
Extremadura	5
Galicia	3
Madrid, Comunidad de	6
Murcia, Región de	5
Navarra, Comunidad Foral	2
País Vasco	6
Rioja, La	2

*Fuente: Elaboración propia*

En la tabla 3.15 podemos observar como de nuevo existen dos clústeres individuales, Andalucía (1), como en el anterior caso, e Islas Baleares (4). Además, Navarra ahora forma parte del clúster 2.

Estas discrepancias pueden verse observadas en la tabla 3.8, donde Andalucía sigue teniendo una mala situación económica, pero con elevada población e industrialización y destaca por su población no envejecida. El otro conglomerado con mala situación económica, donde podría pertenecer Andalucía, es el (5), pero el conglomerado 5 tiene poca población, por lo que no coincide con las características de Andalucía. Por lo tanto, vuelve a formar parte como un conglomerado individual.

Al mismo tiempo, Islas Baleares destaca por tener la población más joven. Esto tiene sentido ya que en el análisis descriptivo vimos como esta Comunidad es en la que más crece la población y eso indica que su población es joven. Los demás conglomerados con población joven son el (1) y (6) pero no puede formar parte de ninguno de estos debido a que tienen muchas diferencias en los demás componentes.

En ambos casos estas diferencias significativas permiten separar a ambas

comunidades en un conglomerado individual para Islas Baleares y otro para Andalucía.

**Tabla 3.16 Resumen De Casos**

Cluster		Buena Situación Económica	Población No Envejecida y Alto Precio Vivienda	Industrialización y Elevado Población
1	Media	-1,39	0,57	1,77
2	Media	0,64	-0,15	-1,01
3	Media	0,08	-1,60	0,38
4	Media	0,08	1,86	-1,52
5	Media	-1,08	0,20	-0,16
6	Media	1,30	0,65	1,16

*Fuente: Elaboración propia*

El clúster 1 está formado únicamente por Andalucía al igual que en el caso de 2010. Implica una mala situación económica con elevada población e industrialización donde la población no está envejecida. Si lo comparamos con el caso 2010 este clúster formado por esta única Comunidad Autónoma tiene las mismas características. La peor situación económica la tiene Andalucía, lo que nos puede indicar que es probablemente la Comunidad que más se ha visto afectada por la crisis, en términos económicos.

El clúster 2 incluye a Aragón, Cantabria, Navarra y La Rioja. Indica una buena situación económica con poca población e industrialización donde la población está algo envejecida.

En este clúster encontramos la primera diferencia respecto a 2010, ya que Navarra estaba en un clúster independiente y Cantabria también pertenecía a otro.

Cantabria se diferenciaba en 2010, de La Rioja y Aragón en cuanto a población envejecida. Aunque ambos clústeres implican población envejecida, en Cantabria hay mayor envejecimiento tanto en 2010 como en 2015. Cantabria ha estado ligeramente por encima de esas dos Comunidades en “edad media de la población” (Tablas del Anexo A.9, A.10). Es en 2015, donde La Rioja y Aragón aumentan su población envejecida y se acercan a la de Cantabria, por lo que se

juntan en este conglomerado por este atributo.

En el caso de Navarra se agrupa al conglomerado dos por el componente de industrialización y población. Esta Comunidad destaca por sus bajos valores en ambos componentes. Como ya hemos comentado en el primer apartado de este trabajo, el Índice industrial se reduce en 2015, en comparación con 2010. La Rioja y Aragón han perdido IPI durante estos años, lo que ha permitido igualar estas Comunidades y agruparlas en un mismo clúster.

El clúster 3 está constituido por Asturias, Castilla y León, Galicia con una situación económica estándar, al igual que población e industrialización, pero con la población más envejecida de todas. En comparación con el clúster de 2010 esta clasificación es la misma, con la diferencia de que Cantabria ha pasado al Clúster 2 lo que nos indica que Cantabria no tiene una población tan envejecida como para clasificarla en el Clúster con las Comunidades Autónomas más envejecidas.

Por último, en el gráfico 1.2 vimos como Galicia, Asturias y Castilla y León son justamente las Comunidades que pierden población. Todas a excepción de Extremadura forman parte de este clúster.

El clúster 4 está formado únicamente por Islas Baleares. Esta clasificación nos indica que es una comunidad con una situación económica estándar, pero baja industrialización, con poca población, aunque no envejecida.

El problema que ha ocurrido con Islas Baleares, con respecto a 2010, es su gran disminución del IPI, pasando de 103% a 85%. Esta diferencia y caída del IPI permite obtener diferencias significativas con el clúster en el que estaba en 2010, ya que en Valencia y Murcia el IPI apenas varía. Además, podemos observar una gran diferencia en cuanto a la situación económica en el conglomerado donde antes estaba clasificada, ya que Baleares dentro de ese conglomerado, tenía la mejor situación económica. Ahora, en 2015, esa diferencia es más notable, por ejemplo con Castilla la Mancha y Extremadura. Por tanto, esto nos permite afirmar que Baleares, debido a sus diferencias con el anterior clúster clasificado, empeora su IPI y su mejor situación económica permite separarla en un clúster independiente.

El clúster 5 es el más numeroso, formado por Canarias, Castilla la Mancha,

Comunidad Valenciana, Extremadura y Murcia. Tienen una mala situación económica, baja industrialización y población, pero su población no está envejecida. Hay que recordar que, en 2010, este conglomerado tenía la peor situación económica.

En 2010, Baleares estaba incluida en este clúster, pero ahora esta Comunidad forma parte de un clúster unitario por las altas diferencias con este conglomerado como ya hemos comentado.

El clúster 6 incluye al País Vasco, Madrid y Cataluña. Este conglomerado continúa siendo el clúster dominante en prosperidad, al igual que en 2010. Existe una buena situación económica, industrialización, elevada población y no está envejecida.

En la tabla 3.17 pretendo demostrar si son adecuados los Clúster que están formados por una sola Comunidad.

**Tabla 3.17 Índice De Silueta**

<b>Caso</b>	<b>Clúster</b>	<b>SM6</b>	<b>S6</b>
<b>Andalucía</b>	1	5	1
<b>Aragón</b>	2	3	0,668
<b>Asturias, Principado de</b>	3	2	0,857
<b>Baleares, Illes</b>	4	2	1
<b>Canarias</b>	5	1	0,606
<b>Cantabria</b>	2	3	0,669
<b>Castilla y León</b>	3	2	0,903
<b>Castilla - La Mancha</b>	5	2	0,653
<b>Cataluña</b>	6	1	0,696
<b>Comunitat Valenciana</b>	5	1	0,43
<b>Extremadura</b>	5	2	0,582
<b>Galicia</b>	3	2	0,895
<b>Madrid, Comunidad de</b>	6	1	0,69
<b>Murcia, Región de</b>	5	4	0,558
<b>Navarra, Comunidad Foral</b>	2	4	0,45
<b>País Vasco</b>	6	2	0,342
<b>Rioja, La</b>	2	5	0,659

*Fuente: Elaboración propia*

Como existen dos conglomerados que solo incluyen un caso, compruebo que todas las CCAA están bien clasificadas, no existiendo ninguna con S6 negativo que explicaría que está mal clasificada.

En este caso es interesante fijarse en las Islas Baleares y en Andalucía, según SM6 (siguiente mejor clúster) serían el 2 y el 5 respectivamente. A continuación, voy a considerar que están en esos clústeres para ver si estarían correctamente clasificadas o no.

Si Islas Baleares estuviera clasificado en el clúster 2 entraría en contradicción con el factor de población no envejecida ya que aquí la población estaría algo envejecida y realmente en 2015 la población de las Islas Baleares es de las más jóvenes ya que su población edad media es muy baja. Por tanto, si Islas Baleares estuviera en el clúster 2, estaría mal clasificada por el componente población no envejecida.

Si Andalucía estuviera clasificada en el clúster 5 entraría en contradicción en el factor de población e industrialización porque en este clúster el valor de este factor es bajo. Además, en 2015, Andalucía tiene el número de habitantes más alto y un índice de industrialización por encima de la media, por lo tanto, estaría erróneamente clasificada.

Finalmente, puedo concluir que es óptima la clasificación en seis clústeres porque si no habría una mala clasificación de las Comunidades Autónomas.

### 3.4 Caso 3: Año 2021

El año 2021 ha sido elegido como año de estudio ya que la pandemia Covid-19 suponemos que ha afectado a las distintas Comunidades. Además, es el último año para el que tenemos datos oficiales, no provisionales.

**Tabla 3.18 Pertinencia Del Análisis**

Medida Kaiser-Meyer-Olkin	KMO	0,517
	Aprox. Chi-cuadrado	328,75
Prueba de esfericidad de Bartlett	gl	91
	Sig.	0

*Fuente: Elaboración propia*

Como ya he comentado en los anteriores casos, el análisis factorial exploratorio con este conjunto de datos perteneciente a 2021 es posible realizarlo, ya que, la pertinencia del análisis permite aceptar los dos criterios vistos con anterioridad. La medida de Kaiser (KMO) es superior a 0,5 y la prueba de esfericidad de Bartlett nos permite afirmar la presencia de factores comunes ya que el p-valor

tiende a cero, por lo que se puede continuar con el análisis.

**Tabla 3.19 Varianza Total Explicada**

Componente	Suma de cargas al cuadrado de la extracción			Suma de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	6,29	44,96	44,96	5,99	42,79	42,79
2	3,44	24,60	69,56	2,55	18,23	61,02
3	1,42	10,12	79,68	2,44	17,46	78,48
4	1,21	8,64	88,32	1,38	9,84	88,32

*Fuente: Elaboración propia*

En los anteriores casos, 2010 y 2015, la varianza total explicada la componían tres factores, ahora la componen cuatro, con autovalores superiores a uno y una varianza total acumulada superior al 70%.

Las varianzas explicadas por cada factor son: 42,79%, 18,23%, 17,46% y 9,84%, respectivamente, sumando una acumulada de 88,32%. Esto implica que se ha perdido muy poca información con el resumen de las variables en factores.

El cuarto factor nos da muy poca información sobre el análisis, tan solo un 9,84%, aun así, su autovalor es 1,21. Aunque con el tercer componente ya llegamos al 70% de varianza acumulada, decido incluirla en el análisis para obtener mejores conclusiones.

**Tabla 3.20 Comunalidades**

Variables	Extracción
S_Vegetativo	0,918
IPC_Índice	0,646
IPI_Índice	0,862
IPV_Índice	0,720
PIBpm	0,976
Población	0,981
Renta_Media_H	0,969
Renta_M_UnidC	0,966
T_Paro	0,853
Riesgo_pobreza	0,956
Edad_Media	0,890
Gasto_M_Persona	0,911
G_I+D_PIB	0,800
C_Laboral_ind	0,916

*Fuente: Elaboración propia*

Podemos observar en esta tabla de comunalidades el porcentaje de varianza de

cada variable lo que explica a los tres factores donde destaca “PIBpm” y “Población”, por el contrario, “IPV\_Índice” y “IPC\_Índice” no tienen unos resultados satisfactorios para el análisis.

En comparación con los anteriores análisis podemos concluir que tanto IPC como IPV no son variables muy relevantes para el análisis, ya que no hay muchas otras variables que estén correlacionadas con estas, como también vimos en el análisis descriptivo.

**Tabla 3.21 Matriz De Componentes Rotados**

Variables	Componente			
	1	2	3	4
<b>S_Vegetativo</b>		0,943		
<b>IPC_Índice</b>				-0,626
<b>IPI_Índice</b>				0,895
<b>IPV_Índice</b>		0,627		
<b>PIBpm</b>			0,929	
<b>Población</b>			0,968	
<b>Renta_Media_H</b>	0,957			
<b>Renta_M_UnidC</b>	0,979			
<b>T_Paro</b>	-0,859			
<b>Riesgo_pobreza</b>	-0,950			
<b>Edad_Media</b>		-0,904		
<b>Gasto_M_Persona</b>	0,930			
<b>G_I+D_PIB</b>	0,772			
<b>C_Laboral_ind</b>	0,873			

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

a La rotación ha convergido en 5 iteraciones.

*Fuente: Elaboración propia*

En este último caso, la matriz de componentes rotados nos ha dado cuatro componentes de los tres que teníamos anteriormente.

El primer factor se agrupa en “**Buena situación Económica**”, donde siguen predominando variables macroeconómicas relacionadas con los gastos, rentas y tasas de paro y de pobreza.

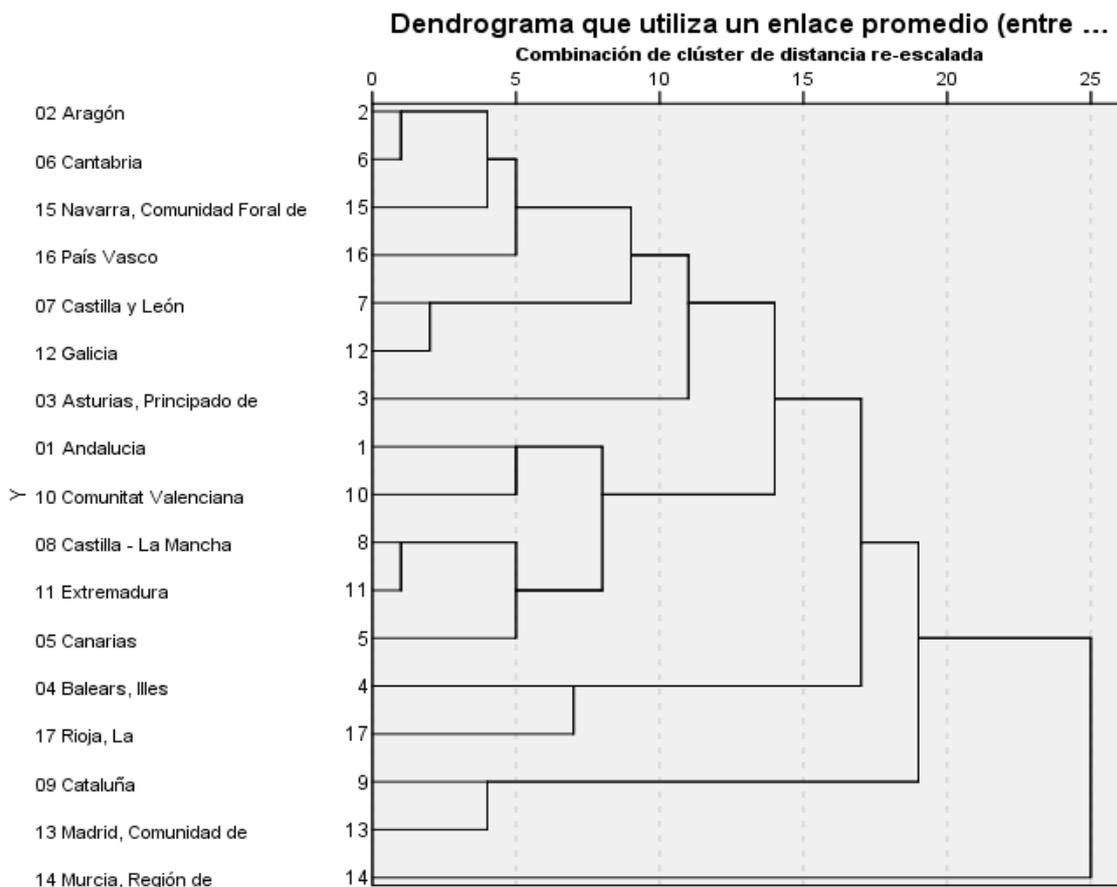
El segundo factor corresponde, como en el caso de 2010, a “Población no envejecida con alto precio de vivienda” aunque el precio de vivienda no es

excesivamente determinante, como hemos observado en Comunalidades así que tan solo lo nombraremos como **“Población no envejecida”**.

En el tercer factor, podemos observar la primera diferencia. En los anteriores casos, nos agrupaba el factor “Elevada población con industrialización”, ahora nos separa industrialización, incluida en el siguiente factor. Por tanto, este factor será **“Elevada población y PIB”**.

En este factor debemos tener en cuenta que, a mayor población, más PIB en términos generales, por eso es interesante tener el PIB per cápita (tabla 1.4 del análisis descriptivo) para obtener interpretaciones más verosímiles.

El cuarto factor está formado por industrialización y el IPC, pero el IPC como hemos visto en la matriz de comunalidades no es muy relevante y además en este factor apenas afecta, entonces para conseguir una mejor interpretación, voy a situar este factor como **“Industrialización”**.



*Fuente:Elaboración propia*

El dendrograma nos facilita que la solución del número de clúster es entre 5 y 6,

según lo observado en la etapa 10.

**Tabla 3.22 Determinación Del Número De Clúster**

Nº Clúster	5	6
Índice Silueta	0,46	0,50
CH	1,9	2,5

*Fuente: Elaboración propia*

El número de clúster óptimo según el índice de silueta y el CH son seis clústeres como en los casos anteriores.

**Tabla 3.23 Caso De Clúster**

Caso	6 clústeres
Andalucía	1
Aragón	2
Asturias, Principado de	3
Balears, Illes	4
Canarias	1
Cantabria	2
Castilla y León	2
Castilla - La Mancha	1
Cataluña	5
Comunitat Valenciana	1
Extremadura	1
Galicia	2
Madrid, Comunidad de	5
Murcia, Región de	6
Navarra, Comunidad Foral	2
País Vasco	2
Rioja, La	4

*Fuente: Elaboración propia*

De nuevo, tenemos dos Comunidades que quedan agrupadas en un clúster individualmente: Asturias (3) y Murcia (6). Ambas Comunidades en 2010 y 2015, estaban agrupadas en otros clústeres y nunca han sido independientes. Esto puede ser una consecuencia de que ahora tengamos cuatro componentes en vez de tres.

El caso de Asturias (3), su cambio con respecto a 2010 y 2015 se debe principalmente al índice de industrialización que ha pasado del 95% en 2015 al 107% en 2021, de las más altas junto a Madrid y Murcia. Asturias antes formaba

parte del conglomerado junto a Castilla y León y Galicia que, a pesar de haber aumentado también su IPI, el incremento no ha sido tan considerable como en el caso de Asturias.

En el caso de Murcia (6), su IPI en 2015 es de 94.7, y de 120 en 2021, siendo la Comunidad con mayor índice en este año. Antes estaba clasificada con otras comunidades como Valencia y Extremadura que, a pesar de tener similitud en los demás componentes, en el caso de la Industrialización hay diferencias muy significativas que permite separarla en un clúster independiente.

Por tanto, ambas Comunidades (Asturias, Murcia) han sido separadas de sus anteriores conglomerados por el componente de Industrialización. Las dos Comunidades han tenido un aumento considerable del IPI, pero no podrían agruparse en un solo clúster, las dos juntas por el componente de envejecimiento de la población: Asturias es de las más ancianas y Murcia de las más jóvenes.

**Tabla 3.24 Resumen De Casos**

<b>Cluster</b>		<b>Buena Situación Económica</b>	<b>Población No Envejecida</b>	<b>Elevada Población y PIB</b>	<b>Industrialización</b>
<b>1</b>	Media	-1,15	0,03	0,38	-0,09
<b>2</b>	Media	0,68	-0,62	-0,30	-0,13
<b>3</b>	Media	0,11	-1,47	-0,65	1,59
<b>4</b>	Media	0,05	1,07	-0,94	-1,54
<b>5</b>	Media	1,12	0,82	1,71	0,16
<b>6</b>	Media	-0,72	1,21	-0,96	2,40

*Fuente: Elaboración propia*

El clúster 1 está formado por Andalucía, Canarias, Comunidad Valenciana, Castilla la Mancha y Extremadura. Tienen la peor situación económica, elevada población y/o PIB, con una industrialización estándar.

Este clúster es muy diferente con respecto a los anteriores casos, ahora Andalucía forma parte de un clúster en conjunto con otras Comunidades.

Por otro lado, el resto de las comunidades ya conformaban en 2015 este clúster, excepto Murcia que ahora está separada. Sigue existiendo mala situación económica con industrialización algo baja y se han producido cambios en cuanto

a la población, ahora más envejecida y hay más número de habitantes. Es, precisamente, esta diferencia de la edad media de la población la que ha permitido separar de este clúster a Murcia, ya que es una de las Comunidades con la población más joven.

El clúster 2 es el conglomerado más numeroso, formado por Aragón, Cantabria, Castilla y León, Galicia, Navarra y País Vasco. Tienen buena situación económica, con PIB y Población baja, industrialización estándar y su población está envejecida.

Este conglomerado ha sufrido muchos cambios. En primer lugar, Navarra, Aragón y Cantabria formaban un conglomerado en conjunto donde destacaban por tener buena situación económica, población algo envejecida y el número de habitantes e industrialización era bajo. En segundo lugar, otro conglomerado estaba formado por Castilla y León y Galicia, donde destacaban principalmente por ser la población más envejecida. Por último, el País Vasco, que siempre ha estado junto a Madrid y Cataluña (Comunidades más prósperas y con todos los componentes en positivo), ahora forma parte de este clúster.

En definitiva, lo sorprendente de este conglomerado es la cercanía geográfica de estas comunidades (forman todas parte del norte de España). Además, este conglomerado se ha formado por tener la población envejecida, pero con una buena situación económica. Lo que nos puede indicar este conglomerado es que, como normal general, la zona norte de España no tiene una mala situación económica, pero sí que tiene una población más envejecida, comparándola con el resto de zonas.

El clúster 3 está compuesto únicamente por Asturias, la cual tiene una situación económica estándar con poca población y/o PIB, alta industrialización y población muy envejecida. Parece que Asturias no forma parte del clúster anterior, por tener la población más envejecida y por estar muy industrializada con respecto al clúster dos.

El clúster 4 está formado por La Rioja e Islas Baleares, con una situación económica estándar, con poco PIB y población, muy poca industrialización y su población no está envejecida.

Islas Baleares ha cambiado de clúster siempre (en los tres casos analizados),

principalmente debido al IPI que ha sufrido enormes cambios en cada periodo analizado: 103.6% (2010), 85% (2015), 78,2% (2021). Por el contrario, La Rioja no ha sufrido cambios tan significativos en sus componentes y se separa de su clúster (que lo formaba con Aragón, Cantabria y Navarra) principalmente por el IPI que en este año es muy bajo.

El clúster 5 incluye a Madrid y Cataluña, siguen siendo las mejores Comunidades con todos los factores positivos, tienen la mejor situación económica, población, PIB e industrialización alta, y su población no está envejecida.

El País Vasco se ha separado de este Clúster dado que su población no es tan numerosa. Además, se ha envejecido la población más que en Madrid y Cataluña.

Madrid y Cataluña siempre han formado parte del mismo clúster con las mismas características positivas, lo que nos permite indicar que son las dos Comunidades más prósperas de España.

El clúster 6 está compuesto únicamente por Murcia, la cual tiene una mala situación económica, PIB y/o población baja, con industrialización muy alta, y su población es la más joven de España.

Este aumento en la producción industrial de Murcia no producido anteriormente se puede deber principalmente a la industria alimentaria, muy importante en Murcia.<sup>24</sup>

En la tabla 3.25 se realiza un análisis para determinar si todas las Comunidades han sido correctamente clasificadas en su respectivo clúster. En caso de que alguna no esté clasificada adecuadamente, se identificará a en qué clúster debería pertenecer y se evaluará si esto mejoraría o empeoraría la interpretación de los resultados.

---

<sup>24</sup> Europa Press. (2020). “La demanda de frutas y verduras se disparó un 40% durante el confinamiento”. El aumento de la demanda de las frutas y verduras durante el confinamiento y el descenso del resto de sectores pueden explicar el gran aumento del IPI de Murcia con respecto a las demás Comunidades

**Tabla 3.25 Índice De Silueta**

<b>Caso</b>	<b>Clúster</b>	<b>SM6</b>	<b>S6</b>
<b>Andalucía</b>	1	5	0,352
<b>Aragón</b>	2	4	0,404
<b>Asturias, Principado de</b>	3	2	1
<b>Balears, Illes</b>	4	1	0,693
<b>Canarias</b>	1	6	0,363
<b>Cantabria</b>	2	3	0,35
<b>Castilla y León</b>	2	1	0,332
<b>Castilla - La Mancha</b>	1	2	0,401
<b>Cataluña</b>	5	2	0,746
<b>Comunitat Valenciana</b>	1	2	0,384
<b>Extremadura</b>	1	4	0,44
<b>Galicia</b>	2	3	0,111
<b>Madrid, Comunidad de</b>	5	2	0,768
<b>Murcia, Región de</b>	6	3	1
<b>Navarra, Comunidad Foral</b>	2	4	0,334
<b>País Vasco</b>	2	5	0,357
<b>Rioja, La</b>	4	2	0,55

*Fuente: Elaboración propia*

Como podemos observar, todas las “S6” son positivas por lo que todas están bien clasificadas.

Por otro lado, es interesante comprobar aquellos clústeres que están formados por solo una Comunidad para ver cómo podrían variar. En el caso de Asturias que forma parte del clúster 3, su siguiente mejor clasificación sería en el clúster 2. La diferencia principal entre estos dos clústeres se encuentra en la industrialización, donde Asturias es muy fuerte en este aspecto. Respecto al componente de envejecimiento, Asturias está por encima de estas comunidades, pero tampoco estaría mal clasificado ya que ambos clústeres están muy envejecidos. Por tanto, Asturias estaría mal clasificada en su siguiente mejor clúster que es el 2 por el componente de Industrialización, por lo que su clasificación en el clúster 3 es correcta.

En el caso de Murcia que forma parte del clúster (6), su siguiente mejor clasificación sería en el clúster 3, es decir, con Asturias. Esto sería un grave error ya que Murcia contiene a la población más joven y Asturias a la población más anciana, por lo que no tendría sentido esta clasificación. Entonces Murcia está bien clasificada en el clúster al que pertenece, el 6.

#### 4. CONCLUSIONES

Con la realización del presente trabajo se ha llegado a algunas conclusiones que pueden ser relevantes. En primer lugar, el análisis descriptivo nos ha permitido observar algunas importantes relaciones entre las variables. Por ejemplo, que el aumento de la tasa de desempleo provoca una disminución de las rentas y gastos medios que se traduce en una caída del PIB y un aumento del riesgo de pobreza. También, comprobamos que la pérdida de población está estrechamente relacionada con el saldo migratorio y el envejecimiento de la población, que ha ido en aumento en los últimos años. Finalmente, este análisis nos ha permitido identificar aquellas variables que no están relacionadas como los índices de precios, de vivienda y consumo, lo que sugiere que estas dos variables pueden ser descartadas sin afectar a los resultados finales. El análisis estadístico posterior, ha ratificado estos resultados.

El análisis factorial, realizado en el tercer apartado del trabajo, nos ha proporcionado el resumen de la información de las variables durante los dos primeros años de estudio, 2010 y 2015, en tres componentes: “buena situación económica”, compuesta por variables macroeconómicas que establecen la riqueza de las familias y del país; “población no envejecida”, formada por el saldo vegetativo y la edad media de la población y, “elevada población e industrialización”, donde destaca el índice de producción industrial y la población. Con respecto a 2021, se producen cambios ya que se añade un componente más, “industrialización”, mejorando las interpretaciones.

El análisis clúster posterior nos brinda la oportunidad de agrupar las Comunidades Autónomas por similitudes y observar cómo estas evolucionan a lo largo del tiempo y cómo se agrupan en nuevos clústeres en función de distintas características. A través de este estudio, creo que hemos podido obtener interpretaciones muy interesantes, como por ejemplo que Madrid y Cataluña son las regiones más prósperas, con todos sus componentes en positivo. País Vasco, que tiene un parecido en algunos componentes con aquellas dos regiones más avanzadas en términos económicos, diverge en cuanto al envejecimiento de la población.

Otro resultado llamativo es que las Comunidades que pertenecen a la zona norte de España, tienen como característica principal poblaciones envejecidas. Podría ser objeto de un estudio posterior, analizar por qué, en general, las regiones del norte presentan un envejecimiento poblacional mayor que las del sur.

Centrándonos en nuestra Comunidad, en Castilla y León, hay que decir que, apenas sufre variaciones durante el período de tiempo analizado. La región destaca por su "despoblación" y población muy envejecida, mientras que, en todos los demás aspectos, como situación económica, industrialización y número de habitantes, es estándar.

Como resultado final, podemos concluir que, aunque las Comunidades Autónomas cambian de clústeres a lo largo de los periodos analizados, las características que definen a cada una de ellas, apenas varían.

## 5. BIBLIOGRAFÍA

- Datos macro (2022), datosmacro.com. (online) Disponible en: <https://datosmacro.expansion.com/>
- De la fuente Fernández, S. (2011), Análisis factorial, (online), Disponible en: <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANT E/FACTORIAL/analisis-factorial.pdf>
- Economipedia (2021). Cuartil. (Online), Disponible en: de <https://economipedia.com/definiciones/cuartil.html?nab=1>
- Europa Press. (2020, 30 de septiembre). “La demanda de frutas y verduras se disparó un 40% durante el confinamiento”. Europa Press (online). Recuperado el [03/03/2023], Disponible en: <https://www.europapress.es/economia/noticia-demanda-frutas-verduras-disparo-40-confinamiento-encima-alimentacion-seca-20200930161951.html>
- Hernández de Cos, Pablo (2018), La recuperación de la economía española, evolución reciente y perspectivas del mercado inmobiliario, (online), Disponible en: <https://www.bde.es/f/webbde/GAP/Secciones/SalaPrensa/IntervencionesPublicas/DirectoresGenerales/economia/Arc/Fic/eco160218.pdf>
- Hernández Vera, Pablo (2020), Fenómeno migratorio en España durante el siglo XXI y sus repercusiones económicas en la actualidad, (online), Disponible en: [https://www.um.es/documents/2918258/18874499/Escrita\\_CCSS\\_IES+Ruiz+de+Alda+-+copia.pdf/d28cbd78-9cfc-4f11-83eb-f001358a6583](https://www.um.es/documents/2918258/18874499/Escrita_CCSS_IES+Ruiz+de+Alda+-+copia.pdf/d28cbd78-9cfc-4f11-83eb-f001358a6583)
- IBM SPSS Statistics 26 (2022), (online) Disponible en: <https://www.ibm.com/es-es/products/spss-statistics/details>
- INE (2023), Instituto Nacional de Estadística. INE. (online) Disponible en: <https://www.ine.es/index.html>
- Instituto Nacional de Estadística (2023), “Índice de Precios Industriales (IPI)”, (Online), disponible en: [https://www.ine.es/prensa/ipi\\_prensa.htm](https://www.ine.es/prensa/ipi_prensa.htm)
- Instituto Nacional de Estadística (2023), “Índice de precios de vivienda (IPV)”,

- (Online), disponible en: [https://www.ine.es/prensa/ipv\\_prensa.htm](https://www.ine.es/prensa/ipv_prensa.htm)
- Instituto Nacional de Estadística (2023), “Población en riesgo de pobreza”, (Online), disponible en: [https://www.ine.es/ss/Satellite?L=es\\_ES&c=INESeccion\\_C&cid=1259925455948&p=%5C&pagename=ProductosYServicios%2FPYSLayout&param1=PYSDetalle&param3=1259924822888](https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259925455948&p=%5C&pagename=ProductosYServicios%2FPYSLayout&param1=PYSDetalle&param3=1259924822888)
  - Laborda, Ángel (2016), *Treinta Años de economía española: una visión general*, (online), Disponible en: [https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS\\_CIE/250art02.pdf](https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_CIE/250art02.pdf)
  - Manríquez Gary, J.C (2022), “Técnicas Estadísticas multivariantes para la investigación en la administración” (pp. 16-17). Editorial: Universidad la Concepción, (online), Disponible en: <http://docencia.udec.cl/wp-content/uploads/2022/03/libreria-tecnicas-estadisticas.pdf>
  - Mercedes Prieto, M. (2021), *Análisis de conglomerados (Clúster)*, (online), Disponible en: [https://cursoanterior5.campusvirtual.uva.es/2021\\_2022/pluginfile.php/2899790/mod\\_resource/content/10/Tema5cluster21\\_22.pdf](https://cursoanterior5.campusvirtual.uva.es/2021_2022/pluginfile.php/2899790/mod_resource/content/10/Tema5cluster21_22.pdf)
  - Mercedes Prieto, M. (2021), *Análisis Factorial*, (online), Disponible en: [https://cursoanterior5.campusvirtual.uva.es/2021\\_2022/pluginfile.php/2899810/mod\\_resource/content/7/tema7factorial21\\_22.pdf](https://cursoanterior5.campusvirtual.uva.es/2021_2022/pluginfile.php/2899810/mod_resource/content/7/tema7factorial21_22.pdf)
  - Montanero Fernández, J. (2008), Capítulo 11: Análisis Factorial, Análisis multivariante. Edita Universidad de Extremadura, Servicio de Publicaciones (online), Disponible en: [http://matematicas.unex.es/~jmf/Archivos/ANALISIS\\_MULTIVARIANTE.pdf](http://matematicas.unex.es/~jmf/Archivos/ANALISIS_MULTIVARIANTE.pdf)
  - Montanero Fernández, J. (2008), Capítulo 12: Análisis Cluster, *Análisis multivariante*. Edita Universidad de Extremadura, Servicio de Publicaciones (online), Disponible en: [http://matematicas.unex.es/~jmf/Archivos/ANALISIS\\_MULTIVARIANTE.pdf](http://matematicas.unex.es/~jmf/Archivos/ANALISIS_MULTIVARIANTE.pdf)
  - Okun, A. M. (1975). *Equality and efficiency: The big tradeoff*. Brookings Institution Press.

- Peña, D. (2002), Capítulo 8: Análisis de conglomerados, *Análisis de datos multivariantes*, (online), Disponible en: [https://eva.fing.edu.uy/pluginfile.php/291198/mod\\_resource/content/1/kupdf.net\\_pentildea-daniel-analisis-de-datos-multivariantes-2002pdf.pdf](https://eva.fing.edu.uy/pluginfile.php/291198/mod_resource/content/1/kupdf.net_pentildea-daniel-analisis-de-datos-multivariantes-2002pdf.pdf)

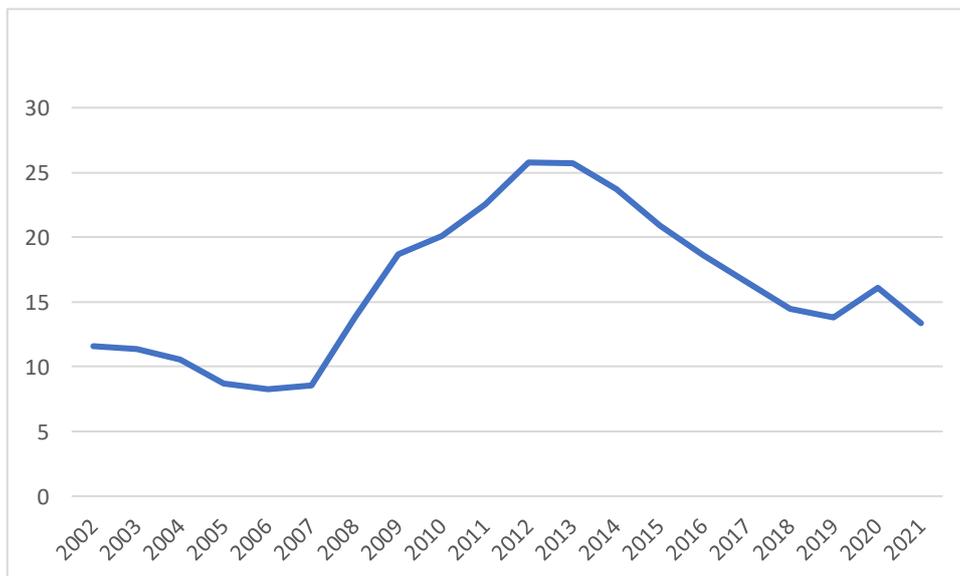
## 6. ANEXOS

**Tabla A.1 Tasa De paro**

Tasa de paro	
2002	11,61
2003	11,37
2004	10,53
2005	8,71
2006	8,26
2007	8,57
2008	13,79
2009	18,66
2010	20,11
2011	22,56
2012	25,77
2013	25,73
2014	23,7
2015	20,9
2016	18,63
2017	16,55
2018	14,45
2019	13,78
2020	16,13
2021	13,33

*Fuente: Elaboración propia a partir de datos del INE*

**Gráfico A.1 Tasa De paro**



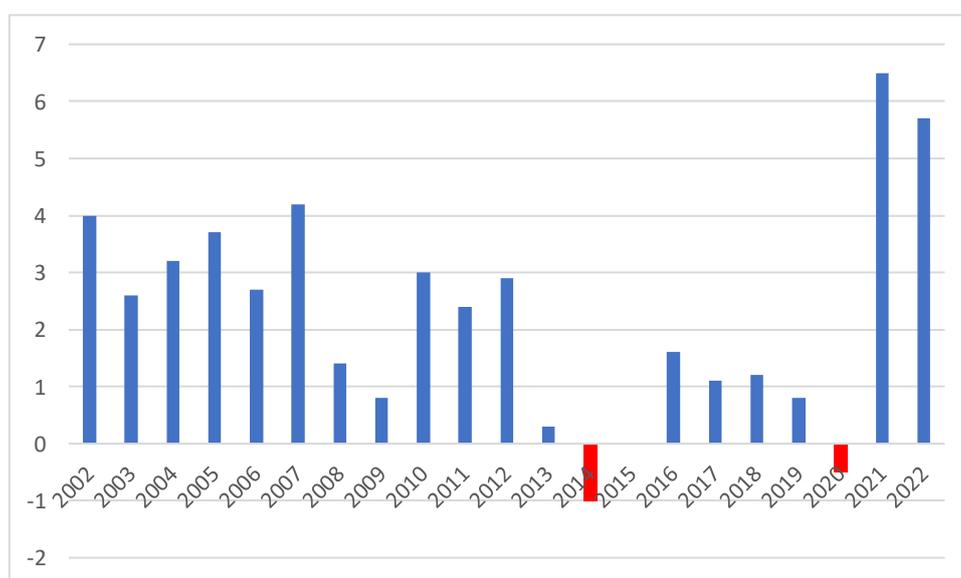
*Fuente: Elaboración propia a partir de datos del INE*

**Tabla A.1 Tasa De Variación Anual Del IPC**

IPC Variación Anual	
2002	4
2003	2,6
2004	3,2
2005	3,7
2006	2,7
2007	4,2
2008	1,4
2009	0,8
2010	3
2011	2,4
2012	2,9
2013	0,3
2014	-1
2015	0
2016	1,6
2017	1,1
2018	1,2
2019	0,8
2020	-0,5
2021	6,5
2022	5,7

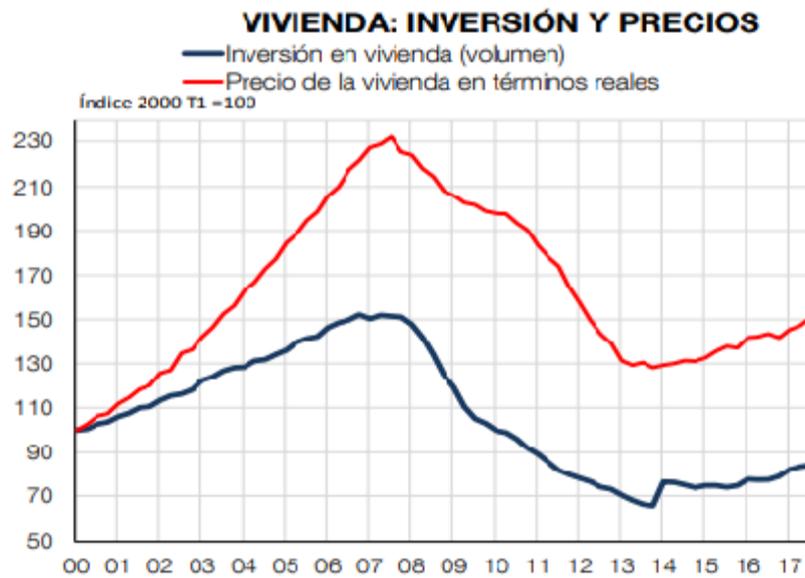
*Fuente: Elaboración propia a partir de datos del INE*

**Gráfico A.1 Tasa De Variación Anual Del IPC**



*Fuente: Elaboración propia a partir de datos del INE*

**Gráfico A.2 Índice De Vivienda: Inversión Y Precios.**



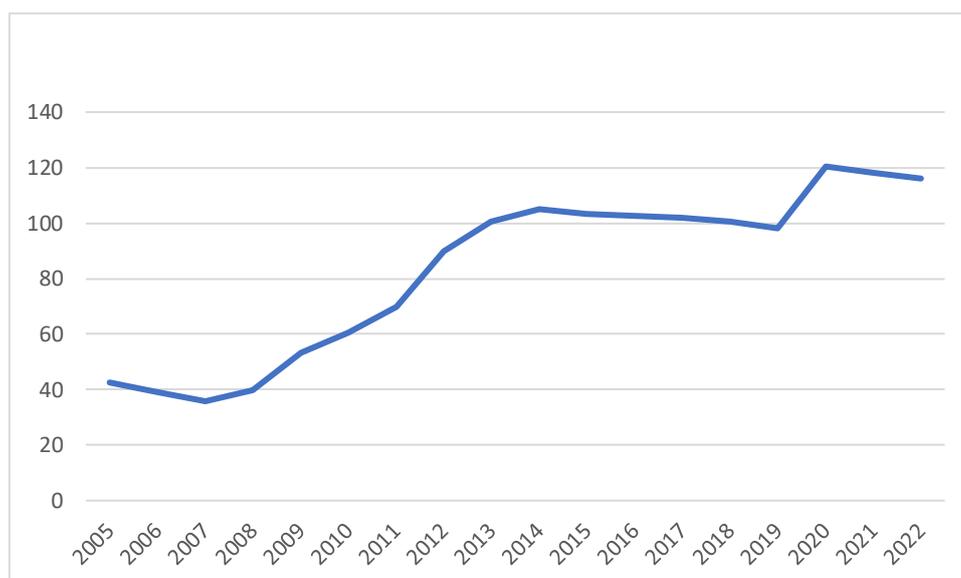
*Fuente: Banco de España, Pablo Hernández 2018.*

**Tabla A.2 Deuda Pública En Porcentaje Del PIB**

Deuda Pública % PIB	
<b>2005</b>	42,4
<b>2006</b>	39,1
<b>2007</b>	35,8
<b>2008</b>	39,7
<b>2009</b>	53,3
<b>2010</b>	60,5
<b>2011</b>	69,9
<b>2012</b>	90
<b>2013</b>	100,5
<b>2014</b>	105,1
<b>2015</b>	103,3
<b>2016</b>	102,7
<b>2017</b>	101,8
<b>2018</b>	100,4
<b>2019</b>	98,2
<b>2020</b>	120,4
<b>2021</b>	118,3
<b>2022</b>	116,1

*Fuente: Elaboración propia a partir de datos del INE*

**Gráfico A.3 Deuda Pública En Porcentaje Del PIB**



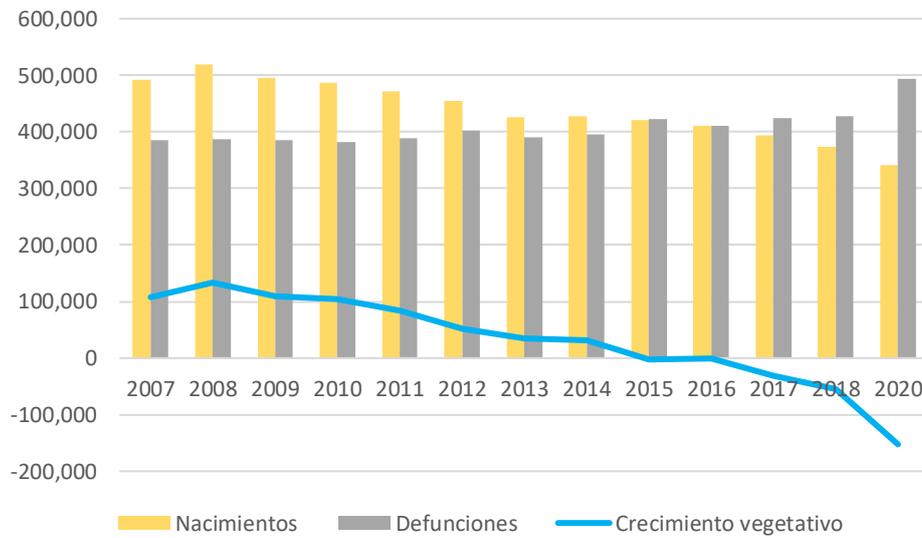
*Fuente: Elaboración propia a partir de datos del INE*

**Tabla A.3 Crecimiento Vegetativo: Natalidad Y Mortalidad**

<b>Años</b>	<b>Nacimientos</b>	<b>Defunciones</b>	<b>Crecimiento vegetativo</b>
2007	492.527	385.361	107.166
2008	519.779	386.324	133.455
2009	494.997	384.933	110.064
2010	486.575	382.047	104.528
2011	471.999	387.911	84.088
2012	454.648	402.950	51.698
2013	425.715	390.419	35.296
2014	427.595	395.830	31.765
2015	420.290	422.568	-2.278
2016	410.583	410.611	-28
2017	393.181	424.523	-31.342
2018	372.777	427.721	-54.944
2020	341.315	493.776	-152.461

*Fuente: Elaboración propia a partir de datos del INE*

### Gráfico A.4 Crecimiento Vegetativo: Natalidad Y Mortalidad



Fuente: Elaboración propia a partir de datos del INE

### Tabla A.4 Tasa De Natalidad: española Y Extranjera

Años	Española	Extranjera
2002	9,51	21,02
2003	9,8	19,99
2004	9,86	19,74
2005	9,91	18,57
2006	10,02	18,86
2007	9,86	19,27
2008	10,1	20,34
2009	9,58	18,69
2010	9,38	18,46
2011	9,16	17,24
2012	8,82	16,78
2013	8,3	16,05
2014	8,37	16,59
2015	8,21	16,71
2016	7,95	16,97
2017	7,53	16,75
2018	7,01	16,27
2019	6,65	15,72
2020	6,28	14,39
2021	6,31	13,51

Fuente: Elaboración propia a partir de datos del INE

Tabla A.5 Correlaciones Pearson 2010

Correlaciones															
		S_Vegetativo	IPC_Índice	IPI_Índice	IPV_Índice	PIBpm	Población	Renta_Media_H	Renta_M_UnidC	T_Paro	T_Riesgo_pobreza	Edad_Media	Gasto_M_Persona	G_I+D_PIB	C_Laboral_in d
S_Vegetativo	Correlación de Pearson	1	-0,134	0,119	-0,351	0,393	0,358	0,133	0,003	0,348	0,247	-,934**	-0,031	0,158	0,127
	Sig.		0,607	0,648	0,167	0,119	0,159	0,610	0,990	0,171	0,340	0,000	0,906	0,545	0,626
IPC_Índice	Correlación de Pearson	-0,134	1	0,182	0,130	-0,341	-0,201	-,519 <sup>†</sup>	-,536 <sup>†</sup>	0,443	,488 <sup>†</sup>	-0,077	-,582 <sup>†</sup>	-0,333	-,583 <sup>†</sup>
	Sig.	0,607		0,484	0,618	0,181	0,438	0,033	0,026	0,075	0,047	0,768	0,014	0,191	0,014
IPI_Índice	Correlación de Pearson	0,119	0,182	1	-,496 <sup>†</sup>	0,256	0,283	-0,280	-0,313	0,338	0,363	-0,194	-0,218	-0,172	0,033
	Sig.	0,648	0,484		0,043	0,321	0,271	0,276	0,221	0,184	0,152	0,456	0,401	0,509	0,901
IPV_Índice	Correlación de Pearson	-0,351	0,130	-,496 <sup>†</sup>	1	-0,294	-0,391	0,255	0,293	-,572 <sup>†</sup>	-0,380	0,476	0,213	0,293	0,054
	Sig.	0,167	0,618	0,043		0,252	0,120	0,324	0,254	0,017	0,132	0,053	0,412	0,254	0,838
PIBpm	Correlación de Pearson	0,393	-0,341	0,256	-0,294	1	,947**	0,323	0,266	0,075	-0,189	-0,285	0,260	0,480	,513 <sup>†</sup>
	Sig.	0,119	0,181	0,321	0,252		0,000	0,206	0,302	0,775	0,468	0,267	0,314	0,051	0,035
Población	Correlación de Pearson	0,358	-0,201	0,283	-0,391	,947**	1	0,111	0,054	0,275	-0,013	-0,327	0,061	0,324	0,312
	Sig.	0,159	0,438	0,271	0,120	0,000		0,671	0,836	0,286	0,961	0,200	0,816	0,205	0,223
Renta_Media_H	Correlación de Pearson	0,133	-,519 <sup>†</sup>	-0,280	0,255	0,323	0,111	1	,984**	-,747**	-,863**	0,169	,925**	,836**	,904**
	Sig.	0,610	0,033	0,276	0,324	0,206	0,671		0,000	0,001	0,000	0,516	0,000	0,000	0,000
Renta_M_UnidC	Correlación de Pearson	0,003	-,536 <sup>†</sup>	-0,313	0,293	0,266	0,054	,984**	1	-,791**	-,902**	0,291	,949**	,801**	,891**
	Sig.	0,990	0,026	0,221	0,254	0,302	0,836	0,000		0,000	0,000	0,256	0,000	0,000	0,000
T_Paro	Correlación de Pearson	0,348	0,443	0,338	-,572 <sup>†</sup>	0,075	0,275	-,747**	-,791**	1	,843**	-,649**	-,798**	-,639**	-,661**
	Sig.	0,171	0,075	0,184	0,017	0,775	0,286	0,001	0,000		0,000	0,005	0,000	0,006	0,004
T_Riesgo_pobreza	Correlación de Pearson	0,247	,488 <sup>†</sup>	0,363	-0,380	-0,189	-0,013	-,863**	-,902**	,843**	1	-,524 <sup>†</sup>	-,887**	-,672**	-,822**
	Sig.	0,340	0,047	0,152	0,132	0,468	0,961	0,000	0,000	0,000		0,031	0,000	0,003	0,000
Edad_Media	Correlación de Pearson	-,934**	-0,077	-0,194	0,476	-0,285	-0,327	0,169	0,291	-,649**	-,524 <sup>†</sup>	1	0,326	0,128	0,165
	Sig.	0,000	0,768	0,456	0,053	0,267	0,200	0,516	0,256	0,005	0,031		0,201	0,624	0,527
Gasto_M_Persona	Correlación de Pearson	-0,031	-,582 <sup>†</sup>	-0,218	0,213	0,260	0,061	,925**	,949**	-,798**	-,887**	0,326	1	,778**	,898**
	Sig.	0,906	0,014	0,401	0,412	0,314	0,816	0,000	0,000	0,000	0,000	0,201		0,000	0,000
G_I+D_PIB	Correlación de Pearson	0,158	-0,333	-0,172	0,293	0,480	0,324	,836**	,801**	-,639**	-,672**	0,128	,778**	1	,805**
	Sig.	0,545	0,191	0,509	0,254	0,051	0,205	0,000	0,000	0,006	0,003	0,624	0,000		0,000
C_Laboral_in d	Correlación de Pearson	0,127	-,583 <sup>†</sup>	0,033	0,054	,513 <sup>†</sup>	0,312	,904**	,891**	-,661**	-,822**	0,165	,898**	,805**	1
	Sig.	0,626	0,014	0,901	0,838	0,035	0,223	0,000	0,000	0,004	0,000	0,527	0,000	0,000	

Fuente: Elaboración propia

**Tabla A.6 Correlaciones Pearson 2015**

		Correlaciones													
		S_Vegetativo	IPC_Índice	IPI_Índice	IPV_Índice	PIBpm	Población	Renta_media_H	Renta_M_UnidC	T_Paro	T_R_Pobreza	Edad_Media	Gasto_M_Persona	G_I+d_PIB	C_Laboral_in d
S_Vegetativo	Correlación de Pearson	1	-,530*	0,125	,632**	0,405	0,366	0,061	-0,051	0,123	0,274	-,944**	-0,058	0,195	0,010
	Sig.		0,029	0,631	0,006	0,107	0,148	0,815	0,845	0,638	0,287	0,000	0,826	0,453	0,969
IPC_Índice	Correlación de Pearson	-,530*	1	-0,224	-0,422	-0,479	-0,422	-0,253	-0,173	-0,129	-0,035	,500*	-0,163	-0,160	-0,297
	Sig.	0,029		0,387	0,092	0,052	0,092	0,327	0,507	0,620	0,893	0,041	0,531	0,541	0,247
IPI_Índice	Correlación de Pearson	0,125	-0,224	1	0,161	,615**	,632**	-0,132	-0,136	0,297	0,267	-0,074	-0,071	0,246	0,268
	Sig.	0,631	0,387		0,537	0,009	0,007	0,613	0,602	0,248	0,301	0,776	0,786	0,341	0,298
IPV_Índice	Correlación de Pearson	,632**	-0,422	0,161	1	,602*	,488*	0,258	0,198	-0,161	-0,080	-,515*	0,294	0,213	0,270
	Sig.	0,006	0,092	0,537		0,011	0,047	0,316	0,445	0,538	0,761	0,035	0,251	0,411	0,295
PIBpm	Correlación de Pearson	0,405	-0,479	,615**	,602*	1	,942**	0,223	0,179	0,055	-0,021	-0,284	0,196	,541*	,493*
	Sig.	0,107	0,052	0,009	0,011		0,000	0,389	0,491	0,833	0,937	0,270	0,451	0,025	0,044
Población	Correlación de Pearson	0,366	-0,422	,632**	,488*	,942**	1	-0,019	-0,059	0,279	0,219	-0,322	-0,032	0,393	0,291
	Sig.	0,148	0,092	0,007	0,047	0,000		0,944	0,822	0,278	0,398	0,207	0,903	0,119	0,257
Renta_media_H	Correlación de Pearson	0,061	-0,253	-0,132	0,258	0,223	-0,019	1	,991**	-,857**	-,901**	0,217	,917**	,718**	,822**
	Sig.	0,815	0,327	0,613	0,316	0,389	0,944		0,000	0,000	0,000	0,404	0,000	0,001	0,000
Renta_M_UnidC	Correlación de Pearson	-0,051	-0,173	-0,136	0,198	0,179	-0,059	,991**	1	-,878**	-,935**	0,324	,928**	,701**	,816**
	Sig.	0,845	0,507	0,602	0,445	0,491	0,822	0,000		0,000	0,000	0,204	0,000	0,002	0,000
T_Paro	Correlación de Pearson	0,123	-0,129	0,297	-0,161	0,055	0,279	-,857**	-,878**	1	,916**	-0,395	-,843**	-,506*	-,611**
	Sig.	0,638	0,620	0,248	0,538	0,833	0,278	0,000	0,000		0,000	0,116	0,000	0,038	0,009
T_R_Pobreza	Correlación de Pearson	0,274	-0,035	0,267	-0,080	-0,021	0,219	-,901**	-,935**	,916**	1	-,525*	-,892**	-,580*	-,736**
	Sig.	0,287	0,893	0,301	0,761	0,937	0,398	0,000	0,000	0,000		0,030	0,000	0,015	0,001
Edad_Media	Correlación de Pearson	-,944**	,500*	-0,074	-,515*	-0,284	-0,322	0,217	0,324	-0,395	-,525*	1	0,338	0,077	0,252
	Sig.	0,000	0,041	0,776	0,035	0,270	0,207	0,404	0,204	0,116	0,030		0,185	0,768	0,329
Gasto_M_Persona	Correlación de Pearson	-0,058	-0,163	-0,071	0,294	0,196	-0,032	,917**	,928**	-,843**	-,892**	0,338	1	,704**	,877**
	Sig.	0,826	0,531	0,786	0,251	0,451	0,903	0,000	0,000	0,000	0,000	0,185		0,002	0,000
G_I+d_PIB	Correlación de Pearson	0,195	-0,160	0,246	0,213	,541*	0,393	,718**	,701**	-,506*	-,580*	0,077	,704**	1	,807**
	Sig.	0,453	0,541	0,341	0,411	0,025	0,119	0,001	0,002	0,038	0,015	0,768	0,002		0,000
C_Laboral_in d	Correlación de Pearson	0,010	-0,297	0,268	0,270	,493*	0,291	,822**	,816**	-,611**	-,736**	0,252	,877**	,807**	1
	Sig.	0,969	0,247	0,298	0,295	0,044	0,257	0,000	0,000	0,009	0,001	0,329	0,000	0,000	

Fuente: Elaboración propia

Tabla A.7 Correlaciones Pearson 2021

Correlaciones															
		S_Vegetativo	IPC_Indice	IPI_Indice	IPV_Indice	PIBpm	Población	Renta_Media_H	Renta_M_UnidC	T_Paro	T_R_Pobreza	Edad_Media	Gasto_M_Persona	G_I+d_PIB	C_Laboral_in d
S_Vegetativo	Correlación de Pearson	1	-0,344	0,135	,543 <sup>ˆ</sup>	0,353	0,322	0,252	0,057	0,246	0,040	-,965 <sup>ˆˆ</sup>	0,003	0,126	0,045
	Sig.		0,177	0,606	0,024	0,164	0,208	0,329	0,828	0,340	0,880	0,000	0,990	0,631	0,865
IPC_Indice	Correlación de Pearson	-0,344	1	-0,318	-0,316	-0,119	-0,100	0,047	0,138	-0,255	-0,243	0,344	0,081	0,110	-0,082
	Sig.	0,177		0,213	0,217	0,650	0,704	0,859	0,598	0,324	0,347	0,177	0,759	0,673	0,755
IPI_Indice	Correlación de Pearson	0,135	-0,318	1	0,003	0,272	0,264	0,186	0,128	-0,250	-0,013	-0,087	0,088	0,361	0,355
	Sig.	0,606	0,213		0,990	0,291	0,305	0,474	0,625	0,333	0,962	0,739	0,738	0,155	0,162
IPV_Indice	Correlación de Pearson	,543 <sup>ˆ</sup>	-0,316	0,003	1	,643 <sup>ˆˆ</sup>	,518 <sup>ˆ</sup>	0,444	0,341	-0,087	-0,331	-0,414	0,440	0,193	0,453
	Sig.	0,024	0,217	0,990		0,005	0,033	0,075	0,180	0,741	0,194	0,098	0,077	0,458	0,068
PIBpm	Correlación de Pearson	0,353	-0,119	0,272	,643 <sup>ˆˆ</sup>	1	,945 <sup>ˆˆ</sup>	0,357	0,286	-0,013	-0,050	-0,317	0,377	,518 <sup>ˆ</sup>	,502 <sup>ˆ</sup>
	Sig.	0,164	0,650	0,291	0,005		0,000	0,159	0,267	0,960	0,848	0,215	0,136	0,033	0,040
Población	Correlación de Pearson	0,322	-0,100	0,264	,518 <sup>ˆ</sup>	,945 <sup>ˆˆ</sup>	1	0,128	0,058	0,212	0,176	-0,343	0,169	0,370	0,278
	Sig.	0,208	0,704	0,305	0,033	0,000		0,624	0,825	0,414	0,500	0,178	0,516	0,144	0,280
Renta_Media_H	Correlación de Pearson	0,252	0,047	0,186	0,444	0,357	0,128	1	,979 <sup>ˆˆ</sup>	-,721 <sup>ˆˆ</sup>	-,882 <sup>ˆˆ</sup>	-0,039	,886 <sup>ˆˆ</sup>	,806 <sup>ˆˆ</sup>	,856 <sup>ˆˆ</sup>
	Sig.	0,329	0,859	0,474	0,075	0,159	0,624		0,000	0,001	0,000	0,883	0,000	0,000	0,000
Renta_M_UnidC	Correlación de Pearson	0,057	0,138	0,128	0,341	0,286	0,058	,979 <sup>ˆˆ</sup>	1	-,786 <sup>ˆˆ</sup>	-,922 <sup>ˆˆ</sup>	0,152	,912 <sup>ˆˆ</sup>	,799 <sup>ˆˆ</sup>	,856 <sup>ˆˆ</sup>
	Sig.	0,828	0,598	0,625	0,180	0,267	0,825	0,000		0,000	0,000	0,561	0,000	0,000	0,000
T_Paro	Correlación de Pearson	0,246	-0,255	-0,250	-0,087	-0,013	0,212	-,721 <sup>ˆˆ</sup>	-,786 <sup>ˆˆ</sup>	1	,851 <sup>ˆˆ</sup>	-0,416	-,711 <sup>ˆˆ</sup>	-,561 <sup>ˆ</sup>	-,704 <sup>ˆˆ</sup>
	Sig.	0,340	0,324	0,333	0,741	0,960	0,414	0,001	0,000		0,000	0,097	0,001	0,019	0,002
T_R_Pobreza	Correlación de Pearson	0,040	-0,243	-0,013	-0,331	-0,050	0,176	-,882 <sup>ˆˆ</sup>	-,922 <sup>ˆˆ</sup>	,851 <sup>ˆˆ</sup>	1	-0,263	-,846 <sup>ˆˆ</sup>	-,578 <sup>ˆ</sup>	-,708 <sup>ˆˆ</sup>
	Sig.	0,880	0,347	0,962	0,194	0,848	0,500	0,000	0,000	0,000		0,309	0,000	0,015	0,001
Edad_Media	Correlación de Pearson	-,965 <sup>ˆˆ</sup>	0,344	-0,087	-0,414	-0,317	-0,343	-0,039	0,152	-0,416	-0,263	1	0,212	0,027	0,146
	Sig.	0,000	0,177	0,739	0,098	0,215	0,178	0,883	0,561	0,097	0,309		0,414	0,919	0,576
Gasto_M_Persona	Correlación de Pearson	0,003	0,081	0,088	0,440	0,377	0,169	,886 <sup>ˆˆ</sup>	,912 <sup>ˆˆ</sup>	-,711 <sup>ˆˆ</sup>	-,846 <sup>ˆˆ</sup>	0,212	1	,802 <sup>ˆˆ</sup>	,875 <sup>ˆˆ</sup>
	Sig.	0,990	0,759	0,738	0,077	0,136	0,516	0,000	0,000	0,001	0,000	0,414		0,000	0,000
G_I+d_PIB	Correlación de Pearson	0,126	0,110	0,361	0,193	,518 <sup>ˆ</sup>	0,370	,806 <sup>ˆˆ</sup>	,799 <sup>ˆˆ</sup>	-,561 <sup>ˆ</sup>	-,578 <sup>ˆ</sup>	0,027	,802 <sup>ˆˆ</sup>	1	,829 <sup>ˆˆ</sup>
	Sig.	0,631	0,673	0,155	0,458	0,033	0,144	0,000	0,000	0,019	0,015	0,919	0,000		0,000
C_Laboral_in d	Correlación de Pearson	0,045	-0,082	0,355	0,453	,502 <sup>ˆ</sup>	0,278	,856 <sup>ˆˆ</sup>	,856 <sup>ˆˆ</sup>	-,704 <sup>ˆˆ</sup>	-,708 <sup>ˆˆ</sup>	0,146	,875 <sup>ˆˆ</sup>	,829 <sup>ˆˆ</sup>	1
	Sig.	0,865	0,755	0,162	0,068	0,040	0,280	0,000	0,000	0,002	0,001	0,576	0,000	0,000	

Fuente: Elaboración propia

### Tabla A.8 Conjunto De Datos 2010

CCAA	S_Vegetativo	IPC_Índice	IPI_Índice	IPV_Índice	PIBpm	Población	R_M_H	R_M_UC	T_Paro	T_R_Pobrezaza	Edad_Media	Gasto_M_Persona	G_IdPIB	C_Laboral_Ind	Buena situación económica	Industrialización y elevada población	Población no envejecida	CLU6_1	SM6	S6
01 Andalucía	3,3	96,8	112,0	126,9	144752225	8370975	25543	14152	28,0	28,3	39,2	10199,4	1,2	2799,1	-0,6	2,0	0,5	1,0	4,0	1,0
02 Aragón	-0,2	96,7	94,4	146,8	33828995	1347095	30166	17773	16,3	14,2	42,9	10985,8	1,1	3016,3	0,2	-0,9	-0,4	2,0	3,0	0,7
03 Asturias, Principado de	-4,6	96,6	99,4	142,2	22733586	1084341	30226	18377	16,6	11,5	45,8	12690,9	1,1	3053,4	0,3	-0,3	-1,9	3,0	2,0	0,6
04 Balears, Illes	3,9	95,0	103,6	123,2	26535097	1106049	29405	17076	21,8	22,7	39,4	11513,6	0,4	2904,6	-0,1	0,0	1,0	4,0	2,0	0,4
05 Canarias	2,7	97,7	109,1	129,5	40811481	2118519	24717	13905	28,8	29,7	39,2	9145,4	0,6	2385,9	-1,5	0,3	0,6	4,0	1,0	0,5
06 Cantabria	0,2	96,0	108,0	132,0	12836685	592250	28237	16456	14,7	20,1	43,0	11948,4	1,2	2976,1	0,0	0,1	-0,7	3,0	2,0	0,4
07 Castilla y León	-2,6	96,2	107,9	144,8	54970431	2559515	27248	16100	15,8	19,5	45,0	10529,8	1,1	2803,7	-0,2	0,4	-1,7	3,0	2,0	0,8
08 Castilla - La Mancha	2,2	97,2	109,0	142,3	38706827	2098373	26717	15130	21,7	28,3	40,6	9862,7	0,7	2555,9	-0,9	-0,1	0,1	4,0	2,0	0,5
09 Cataluña	3,3	94,5	96,1	138,4	201705890	7512381	33297	19239	17,9	14,7	40,8	12308,4	1,7	3295,1	1,4	1,0	0,7	5,0	1,0	0,7
10 Comunitat Valenciana	2,3	96,5	91,0	131,3	101201463	5111706	27057	15780	22,6	22,7	40,8	10545,0	1,1	2592,9	-0,3	0,1	0,7	4,0	2,0	0,5
11 Extremadura	-0,6	97,0	101,1	138,6	18135329	1107220	23349	13509	24,0	35,7	41,8	9277,7	0,9	2250,5	-1,5	-0,4	-0,2	4,0	2,0	0,3
12 Galicia	-2,8	96,5	101,9	135,7	56767263	2797653	27986	15764	15,6	18,1	44,8	10678,6	0,9	2767,8	-0,3	0,4	-1,5	3,0	2,0	0,8
13 Madrid, Comunidad de	5,2	96,3	114,5	134,8	197145571	6458684	35452	20128	15,5	14,4	39,9	13319,5	2,1	3701,2	1,6	1,6	0,6	5,0	1,0	0,6
14 Murcia, Región de	5,5	96,5	102,9	135,5	27912742	1461979	25446	13574	24,5	29,8	38,1	9374,7	0,9	2621,5	-0,9	-0,5	1,5	4,0	2,0	0,4
15 Navarra, Comunidad Foral de	2,8	96,5	86,1	149,1	17973969	636924	37433	21393	11,8	10,7	41,5	13248,8	2,0	3165,4	1,5	-2,0	0,9	6,0	2,0	1,0
16 País Vasco	0,8	95,5	102,7	137,7	64680368	2178339	36211	21484	11,0	11,7	43,4	13596,3	2,0	3510,3	1,6	-0,1	-0,4	5,0	2,0	0,4
17 Rioja, La	1,6	96,4	95,0	152,2	7977333	322415	27104	15769	15,6	21,2	42,0	11265,0	1,1	2561,0	-0,2	-1,5	0,1	2,0	4,0	0,8

*Fuente: Elaboración propia*

Tabla A.9 Conjunto De Datos 2015

CCAA	S_Vegetativo	IPC_índice	IPI_índice	IPV_índice	PIBpm	Población	R_M_H	R_M_UC	T_Paro	T_R_Pobrezaza	Edad_Media	Gasto_M_Persona	G_IdPIB	C_Laboral_ind	Buena situación económica	Población no envejecida con alto precio de viviendas	Industrialización y elevada población	CLU6_1	SM6	S6
01 Andalucía	1,0	100,4	99,2	101,0	144859000	8399043	20851	11862	30	35,7	40,7	9314,4	1,0	2889,9	-1,4	0,6	1,8	1,0	5,0	1,0
02 Aragón	-2,3	100,5	87,6	99,9	32896878	1317847	30058	18395	15	14,4	44,1	11217,3	0,9	3099,4	0,6	-0,4	-0,7	2,0	3,0	0,7
03 Asturias, Principado de	-6,8	100,6	95,5	99,7	21389062	1051229	25909	16227	20	16,7	47,2	11950,6	0,8	3340,9	0,2	-1,9	0,4	3,0	2,0	0,9
04 Balears, Illes	2,0	100,2	85,0	102,3	28271040	1104479	27284	15894	17	21,7	40,7	11704,0	0,3	2766,3	0,1	1,9	-1,5	4,0	2,0	1,0
05 Canarias	0,5	100,4	99,1	100,6	40595518	2100306	22261	12932	27	28,5	41,2	9029,9	0,5	2554,4	-1,2	0,4	0,1	5,0	1,0	0,6
06 Cantabria	-2,9	100,7	84,8	100,9	12336462	585179	25282	15159	18	14,9	44,4	11627,6	0,9	3188,3	0,3	-0,5	-0,9	2,0	3,0	0,7
07 Castilla y León	-4,8	100,8	93,8	100,3	53273013	2472052	24838	15284	18	18,3	46,4	11093,9	1,0	3008,8	0,1	-1,5	0,4	3,0	2,0	0,9
08 Castilla - La Mancha	-0,8	100,8	89,3	100,5	37138498	2059191	21939	12725	25	28,5	42,0	9728,7	0,6	2696,8	-1,0	-0,1	-0,4	5,0	2,0	0,7
09 Cataluña	0,8	100,2	93,5	102,1	204481121	7508106	30655	18236	18	13,9	42,2	12185,6	1,5	3547,2	1,0	1,1	1,3	6,0	1,0	0,7
10 Comunitat Valenciana	-0,2	100,6	93,4	100,9	100185814	4980689	22132	13305	21	25,3	42,4	10201,4	1,0	2604,1	-0,7	0,1	0,5	5,0	1,0	0,4
11 Extremadura	-2,4	100,6	85,9	100,1	17915529	1092997	21092	12421	28	29,0	43,2	9055,3	0,7	2319,0	-1,3	-0,3	-0,7	5,0	2,0	0,6
12 Galicia	-4,5	100,7	94,4	100,0	56697862	2732347	25614	15133	18	19,4	46,2	10570,2	0,9	2951,7	-0,1	-1,5	0,4	3,0	2,0	0,9
13 Madrid, Comunidad de	2,8	100,4	102,0	102,1	204157575	6436996	31243	18593	17	15,1	41,5	12700,0	1,7	3881,8	1,2	1,1	1,7	6,0	1,0	0,7
14 Murcia, Región de	3,2	100,5	94,7	100,4	28508914	1467288	21529	11993	24	31,8	39,7	9620,4	0,9	2713,3	-1,1	0,9	-0,4	5,0	4,0	0,6
15 Navarra, Comunidad Foral de	0,5	100,5	81,5	100,1	18153613	640476	32979	19715	14	9,6	42,7	12677,6	1,6	3370,5	1,4	0,5	-1,3	2,0	4,0	0,4
16 País Vasco	-1,2	100,6	96,2	101,0	65027322	2189257	33053	20264	13	10,9	44,7	14019,7	1,9	3777,2	1,7	-0,3	0,4	6,0	2,0	0,3
17 Rioja, La	-1,2	101,0	87,5	101,2	7966992	317053	26604	16295	14	17,1	43,5	10964,5	0,9	2667,6	0,2	-0,3	-1,2	2,0	5,0	0,7

Fuente: Elaboración propia

Tabla A.10 Conjunto De Datos 2021

CCAA	S_Vegetativo	IPC_Indice	IPI_Indice	IPV_Indice	PIBpm	Población	R_M_H	R_M_UC	T_Paro	T_R_Pobrezaza	Edad_Media	Gasto_M_Persona	G_IdPIB	C_Laboral_ind	Buena situación económica	Población no envejecida	Elevada población y PIB	Industrialización	CLU6_1	SM6	S6
01 Andalucía	-1,6	111,5	100,1	130,2	160747479	8472407	25601	14769	20,2	32,3	42,4	10479,9	1,1	2795,3	-1,5	0,3	1,9	0,0	1,0	5,0	0,4
02 Aragón	-3,9	111,9	96,1	127,0	37964347	1326261	32067	19381	9,0	15,8	45,1	12022,0	1,0	3180,9	0,5	-0,1	-0,6	-0,3	2,0	4,0	0,4
03 Asturias, Principado de	-8,5	110,5	107,1	121,3	23441444	1011792	28369	17968	10,0	20,4	48,8	11855,9	0,9	3501,1	0,1	-1,5	-0,7	1,6	3,0	2,0	1,0
04 Balears, Illes	0,6	111,3	78,2	149,5	30321427	1173008	29368	16867	14,9	17,6	42,0	11954,8	0,5	2943,4	-0,1	2,0	-0,8	-1,6	4,0	1,0	0,7
05 Canarias	-2,0	110,6	90,8	128,6	42655988	2172944	26392	15294	18,9	28,4	43,3	9689,6	0,6	2655,2	-1,2	0,9	-0,4	0,2	1,0	6,0	0,4
06 Cantabria	-4,7	112,0	104,2	127,5	13846024	584507	30447	18659	11,5	15,4	46,1	11270,5	1,0	3145,9	0,3	-0,5	-0,7	0,0	2,0	3,0	0,4
07 Castilla y León	-6,8	112,6	91,3	121,4	58119051	2383139	28890	18181	10,3	17,9	47,9	11930,6	1,4	3021,9	0,2	-1,6	0,3	-0,9	2,0	1,0	0,3
08 Castilla - La Mancha	-2,9	112,7	99,0	113,9	42286077	2049562	26060	15350	13,3	27,4	43,4	9587,5	0,8	2842,8	-0,9	-0,4	-0,1	-0,3	1,0	2,0	0,4
09 Cataluña	-1,5	111,6	97,3	144,7	229418251	7763362	34982	20927	10,2	14,8	43,1	12856,5	1,8	3692,5	0,9	0,5	1,8	-0,4	5,0	2,0	0,7
10 Comunitat Valenciana	-2,8	111,6	106,5	125,4	112603499	5058138	27603	16474	14,4	25,1	43,8	11461,2	1,2	2884,9	-0,6	-0,2	0,8	0,5	1,0	2,0	0,4
11 Extremadura	-4,8	111,6	77,3	111,7	20117062	1059501	22947	13908	18,9	32,3	44,9	10304,3	0,7	2466,3	-1,5	-0,4	-0,3	-0,9	1,0	4,0	0,4
12 Galicia	-6,7	112,1	102,6	122,7	63230043	2695645	27698	16785	11,0	20,2	47,7	11805,5	1,1	3210,1	-0,1	-1,5	0,3	0,1	2,0	3,0	0,1
13 Madrid, Comunidad de	0,2	111,0	109,3	150,3	234639003	6751251	37687	21990	10,1	15,2	42,7	13541,0	1,9	4225,4	1,4	1,1	1,6	0,7	5,0	2,0	0,8
14 Murcia, Región de	0,8	110,4	120,3	123,6	32205157	1518486	27027	14988	12,9	27,7	41,0	10159,2	1,1	2909,4	-0,7	1,2	-1,0	2,4	6,0	3,0	1,0
15 Navarra, Comunidad Foral de	-1,0	111,8	105,4	118,4	20378804	661537	38086	22786	9,9	9,8	43,6	13150,5	1,9	3449,5	1,5	0,5	-1,1	0,4	2,0	4,0	0,3
16 País Vasco	-4,1	111,8	92,9	125,9	71706061	2213993	36887	22658	8,4	12,2	45,8	13981,6	2,3	4163,3	1,7	-0,5	0,1	-0,2	2,0	5,0	0,4
17 Rioja, La	-3,9	112,0	76,2	123,5	8614076	319796	30598	18758	10,5	16,1	44,8	11263,7	0,8	2808,9	0,2	0,1	-1,1	-1,5	4,0	2,0	0,5

Fuente: Elaboración propia

**Tabla A.11 IPC Medio CCAA (2008-2021)**

<b>IPC Medio CCAA desde 2008-2021</b>	
<b>Castilla - La Mancha</b>	101,80
<b>Canarias</b>	101,65
<b>Castilla y León</b>	101,48
<b>Rioja</b>	101,42
<b>Galicia</b>	101,40
<b>Extremadura</b>	101,39
<b>Navarra</b>	101,33
<b>Aragón</b>	101,29
<b>Cantabria</b>	101,29
<b>Andalucía</b>	101,21
<b>Comunidad Valenciana</b>	101,17
<b>Asturias</b>	101,12
<b>Madrid</b>	101,04
<b>Murcia</b>	100,96
<b>País Vasco</b>	100,85
<b>Cataluña</b>	100,56
<b>Baleares</b>	100,51

*Fuente: Elaboración propia a partir de datos del INE*